

Structural and functional approaches to broaden our understanding of coronavirus replication complexes

By:

Thomas K. Anderson

A dissertation in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
(Cellular and Molecular Biology)

at the

University of Wisconsin – Madison
2024

The dissertation is approved by the following members of the Final Oral Committee on 07/12/2024

Thesis Committee:

Robert Kirchdoerfer, Assistant Professor, Biochemistry, Molecular Virology, Quantitative Cell Imaging

Kristen Bernard, Professor, Pathobiological Sciences

Samuel Butcher, Professor, Biochemistry

Nathan Sherer, Professor, Molecular Virology and Oncology

Andrew Mehle, Professor, Medical Microbiology and Immunology

Olve Peersen, Professor, Biochemistry and Molecular Biology

“There is nothing like looking, if you want to find something. You certainly usually find something, if you look, but it is not always quite the something you were after”

J.R.R. Tolkien

Acknowledgements:

To my advisor, Rob Kirchdoerfer, thank you for being an incredible mentor over the past several years. Your scientific rigor, passion, and creativity pushed me to be the best scientist I could be, and I hope to carry this with me throughout both my career and life. When I joined your lab as the first graduate student in 2019, I don't believe either of us knew what we were getting ourselves into, but if I could go back and do it all again, I would.

To my thesis committee members: Dr. Kristen Bernard, Dr. Nathan Sherer, Dr. Samuel Butcher, and Dr. Andrew Mehle. I asked each of you to be on my committee because you are incredible scientists and kind individuals. Your insight, and kindness were so important to my scientific development, and I thank you for your support and guidance.

To the Kirchdoerfer Lab members, thank you for being like a second family through graduate school. The lab has grown immensely these past five years and every individual that has joined brings a unique and fun aspect with them. Having your friendship and support every day made me excited to come to the lab, and I'll miss our constant shenanigans in and out of the workplace.

To Peter and Kennan, thank you for trusting me to play a part in your scientific undergraduate careers. You both played incredibly pivotal roles in helping me throughout my Ph.D., and I cannot thank you enough for that. More importantly, you both taught me countless lessons on how to be a better mentor and for that I will be forever grateful.

To my friends, thank you all for your support and encouragement over the past years. The countless memories and continued joy you provide have been a much-appreciated relief throughout graduate school.

To my CMB cohort, I cannot imagine going through graduate school without each of you. Whether it be trips to the library, bad movie nights, game nights, or potlucks, we were all there for each other to complain about grad school when needed, but more importantly, to celebrate one another's accomplishments.

To my family, thank you for your endless support of me. Mikey, thank you for being a great friend and an even better brother. Charlie and Tara (and kids!), watching you grow as both individuals and a family is incredibly exciting, and I am very grateful I get to be a part of it. To my parents, Chuck and Sue, thank you for always encouraging my wonder and creativity.

To the Sass family, thank you for the countless ways you helped and supported me and Madi over the past many years. I am truly grateful to have had such a supportive family nearby throughout graduate school.

To my wife, Madi, thank you for your never-ending support, encouragement, and love. You are a constant source of joy and silliness that both brightens my life and reminds me not to take myself too seriously. Your passion and drive continuously inspires me to be a better person, friend, and scientist. Thank you for everything.

To Lucy-Loo, ever since you came into our lives you have provided so much happiness and love. Your approach to life reminds me to enjoy the little things and take nothing for granted. Thank you for your support and words of wisdom over the last year.

To Hank the Hedgehog, thank you for being so adorable and instilling a never-ending sense of wonder in me. RIP in peace, king.

Table of contents:

Inspirational quote.....	i
Acknowledgements.....	ii
Table of contents.....	iv
List of figures and tables.....	vii
List of abbreviations.....	xi
Abstract.....	xiii
Chapter 1 – Introduction, The molecular mechanisms of coronavirus replication.....	1
Coronaviruses as human and animal pathogens.....	2
Coronavirus genome and replication.....	5
Structural insights into coronavirus replication.....	15
Coronavirus proofreading.....	20
Coronavirus antivirals.....	23
Concluding remarks	26
References.....	27
Chapter 2 – An alphacoronavirus polymerase structure reveals conserved replication factor functions.....	41
Abstract.....	42
Introduction.....	43
Materials and Methods.....	46
Results.....	55
Discussion.....	68
References.....	74

Supplemental data.....	79
Chapter 3 – A genus-specific nsp12 region impacts polymerase assembly in <i>Alpha</i> - and <i>Gammacoronaviruses</i>	100
Abstract.....	101
Introduction.....	102
Materials and Methods.....	104
Results.....	111
Discussion.....	121
References.....	122
Supplemental data.....	127
Chapter 4 – The coronavirus exonuclease and polymerase interaction is dependent on RNA backtracking.....	141
Abstract.....	142
Introduction.....	143
Methods.....	148
Results.....	154
Discussion.....	163
References.....	167
Supplemental data.....	174
Chapter 5 – Structural insight into araCTP inhibition of the SARS-CoV-2 polymerase complex.....	186
Abstract.....	187
Introduction.....	188

Methods.....	192
Results.....	196
Discussion.....	207
References.....	208
Supplemental data.....	216
Chapter 6 – Discussion and future directions.....	219
Summary.....	220
The coronavirus core-RTC.....	222
Core RTCs from different genera.....	223
Interpreting structures of stalled RTCs.....	225
The CoV RTC in the context of infection.....	226
Coronavirus proofreading and nucleoside analogues.....	227
Interaction between the RTC and exonuclease complex.....	228
Evolution of CoV proofreading.....	231
Concluding remarks.....	232
References.....	232
Appendix 1 – Isolation of a stalled pre-misincorporation SARS-CoV-2 RTC.....	235
Appendix 2 – Development of <i>in vitro</i> RNA assays.....	247
Appendix 3 – In depth protocols used in the completion of thesis work.....	256
Appendix 4 – PDFs of coauthored publications.....	276

List of figures and tables:

Chapter 1 – Introduction, The molecular mechanisms of coronavirus replication

Figure 1, coronavirus taxonomy.....	4
Figure 2, the CoV core-RTC.....	8
Table 1, the CoV nsps.....	12
Figure 3, CoV capping mechanism.....	14
Figure 4, nsp13 mediated backtracking.....	17
Figure 5, theoretical models of CoV proofreading.....	22
Figure 6, CoV antivirals.....	25

Chapter 2 – An alphacoronavirus polymerase structure reveals conserved replication factor functions

Table 1, Native mass spectrometry of coronavirus polymerase complexes.....	56
Figure 1, Assembly of an active PEDV polymerase complex.....	57
Figure 2, Structural differences between PEDV and SARS-CoV-2	60
Figure 3, Effects of replication factor disrupting nsp12 mutations on polymerase activity.....	62
Figure 4, Nsp8 _T 's N-terminal domain is not required for RNA synthesis.....	67
Figure 5, Model of coronavirus backtracking and proofreading.....	71
Table S1, Cryo-EM data collection and refinement.....	79
Table S2, Native mass spectrometry desolvation parameters.....	80
Video S1, 3D variability analysis of PEDV core complex lacking nsp8 _T	80
Figure S1, Cryo-EM validation.....	81
Figure S2, Cryo-EM processing pipeline.....	83

Figure S3, SDS-PAGE of viral proteins.....	85
Figure S4, Native mass spectrometry of coronavirus polymerase complexes.....	86
Figure S5, Full gel images for piecewise and fusion protein primer.....	88
Figure S6, Multiple sequence alignment of coronavirus nsp12s.....	90
Figure S7: Possible cross-effectiveness of CoV antivirals.....	96
Figure S8, Model and density for mutated protein interfaces.....	97
Figure S9, Full gel images for mutant nsp12 primer-extension assays.....	99
Chapter 3 – A genus-specific nsp12 region impacts polymerase assembly in <i>Alpha</i> - and <i>Gammacoronaviruses</i>	
Figure 1, Structure of an active IBV polymerase complex.....	113
Figure 2, Altered sequence and structure of nsp12-nsp8 _F interactions.....	117
Figure 3, Mutant nsp12 primer extensions.....	121
Table S1, Cryo-EM data collection and refinement.....	127
Figure S1 SDS-PAGE analysis of viral RTC proteins.....	128
Figure S2 Cryo-EM data processing pipeline.....	129
Figure S3 Cryo-EM data validation.....	130
Figure S4 Native-PAGE IBV RTC RNA substrate affinity.....	131
Figure S5 IBV RTC Activity Assay.....	131
Figure S6 <i>Coronavirinae</i> nsp8 sequence alignment.....	132
Figure S7 Structural view of IBV specific nsp8 173-181 insert.....	133
Figure S8 Structural view of nsp8 _T 122-129 variable region.....	133
Figure S9 <i>Coronavirinae</i> nsp12 sequence alignment.....	134
Figure S10 Structural view of IBV nsp12 67-72 insertion.....	138

Figure S11 Structural view of IBV nsp12 113-115 shortened loop.....	138
Figure S12 Structural view of IBV nsp12 156-169 insertion.....	139
Figure S13 Mutant nsp12 RTC RNA substrate affinity.....	139
Figure S14 Mutant nsp12 primer extension assays.....	140
Chapter 4 – The coronavirus exonuclease and polymerase interaction is dependent on RNA backtracking	
Figure 1, RTC backtracking.....	145
Table 1, RNA oligo sequences.....	152
Figure 2, Assembly of active SARS-CoV-2 enzyme complexes.....	156
Figure 3, Backtracked RNA promotes proofreading-RTC formation.....	158
Figure 4, NTP channel mutations do not disrupt proofreading-RTC formation...	162
Figure 5, Hypothetical model of CoV proofreading.....	165
Figure S1, Full gel images for SDS-PAGE and activity assays.....	174
Figure S2, Full gel images for piecewise RTC EMSAs.....	175
Figure S3, Mismatched RNA binding of wildtype and D760A-nsp12.....	175
Figure S4, Proofreading-RTC assembly with wildtype nsp14.....	176
Figure S5, Full gel images for backtrack length EMSAS.....	177
Figure S6, Exonuclease complex titration EMSAs.....	178
Figure S7, Full gel images for mutant nsp12 EMSAs.....	179
Figure S8, sequence alignment of coronavirus nsp12s.....	180
Chapter 5 – Structural insight into araCTP inhibition of the SARS-CoV-2 polymerase complex	
Figure 1, AraCTP inhibits the SARS-CoV-2 core-RTC.....	198

Figure 2, Structure of araCTP inhibited core-RTC.....	201
Figure 3, AraCTP in the RdRP active site.....	204
Figure 4, araCMP's altered sugar pucker inhibits RNA elongation.....	206
Table S1, cryo-EM data collection and refinement.....	216
Figure S1, CryoEM Data Processing Pipeline.....	217
Figure S2, CryoEM Validation.....	218
Chapter 6 – Discussion	
Figure 1, Hypothetical model of CoV proofreading.....	230
Appendix 1 – Isolation of a stalled pre-misincorporation SARS-CoV-2 RTC	
Figure 1, Nsp13 reduces <i>in vitro</i> RNA elongation.....	238
Figure 2, Nsp13's helicase activity on the 5' template RNA causes reduction in extension.....	240
Figure 3, Nsp8 _T and nsp13 promote pre-misincorporation pausing.....	242
Appendix 2 – Development of <i>in vitro</i> RNA assays	
Figure 1, <i>in vitro</i> primer extension.....	250
Figure 2, exonuclease assay.....	252
Figure 3, helicase assays.....	254
Appendix 3 – In depth protocols used in the completion of thesis work	
Appendix 4 – PDFs of coauthored publications	

List of abbreviations:

CoV	Coronavirus
SARS-CoV	Severe acute respiratory syndrome coronavirus
SARS-CoV-2	Severe acute respiratory syndrome coronavirus 2
MERS-CoV	Middle East respiratory syndrome coronavirus
COVID-19	Coronavirus induced disease 2019
CCoV-HuPn-2018	Canine coronavirus human pneumonia 2018
PDCoV	Porcine delta coronavirus
IBV	Infectious bronchitis virus
PEDV	Porcine epidemic diarrhea virus
TGEV	Transmissible gastroenteritis virus
SADS-CoV	Swine acute diarrhea syndrome coronavirus
PRCV	Porcine respiratory coronavirus
PHEV	Porcine hemagglutinating encephalomyelitis virus
RNA	Ribonucleic acid
DNA	Deoxyribonucleic acid
ssRNA	Single stranded RNA
dsRNA	Double stranded RNA
polyA	Poly-adenosine
kb	Kilobase
kDa	Kilodalton
nt	Nucleotide
MW	Molecular weight
Nsp	Non-structural protein
RTC	Replication-transcription complex
Core-RTC	Core replication-transcription complex
ORF	Open reading frame
pp	Polyprotein
DMV	Double membrane vesicle
gRNA	genomic RNA
sgRNA	sub-genomic RNA
mRNA	messenger RNA
TRS	Transcriptional regulatory sequence
RdRP	RNA-dependent RNA polymerase
DdDP	DNA-dependent DNA polymerase
NiRAN	Nidovirus RdRP-associated nucleotidyltransferase
NTD	N-terminal domain
Sec	Second
Min	Minute
WT	Wildtype
ER	Endoplasmic reticulum
NTP	Nucleotide triphosphate
GTP	Guanosine triphosphate
ATP	Adenosine triphosphate
CTP	Cytosine triphosphate

UTP	Uridine triphosphate
araNTP	Arabinose nucleotide triphosphate
NA	Nucleoside analogue
Mg ²⁺	Magnesium
NMP	Nucleoside monophosphate
IC50	Inhibitory concentration 50%
RNA Pol II	RNA polymerase II
CryoEM	Cryo electron microscopy
TEM	Transmission electron microscopy
CHAPSO	3-([3-Cholamidopropyl] dimethylammonio)-2-hydroxy-1-propanesulfonate
DDM	n-dodecyl-β-D-maltoside
Nsp8L7	Nsp8 linker nsp7
PDB	Protein data bank
EMDB	Electron microscopy data bank
TEV	Tobacco etch virus
EMSA	Electrophoretic mobility shift assay
5-FU	5-fluorouracil

Abstract:

Coronaviruses (CoVs) are a subfamily of viruses with a propensity to cross over from animal reservoirs into humans, causing epidemics or pandemics. In recent history, there have been three such spillovers: SARS-CoV, MERS-CoV, and SARS-CoV-2, which is the causative agent of COVID-19. CoVs are RNA viruses with uniquely large, ~30 kb (+) single-stranded RNA genomes. Coronaviruses encode over a dozen proteins that are believed to interact and form a viral replication-transcription complex (RTC) that is responsible for the synthesis, capping, and proofreading of viral RNA during an infection. Since the onset of the COVID-19 pandemic, the SARS-CoV-2 RTC has been a focus for biochemical and structural biology studies. The RTC is also a main target for antiviral drug development, with two nucleoside analogue antivirals receiving emergency use authorization for the treatment of COVID-19. Understanding the enzymatic mechanisms by which CoVs replicate and modify their RNA is critical for our ability to develop more antivirals against CoVs to prepare us for future CoV spillovers and diseases.

Most studies on CoV replication have focused on *Betacoronaviruses*, in particular SARS-CoV-2, leaving other genera drastically understudied. Here, we report the first three structures of non-*Betacoronavirus* RTCs, two from the *Alphacoronavirus* genera and one from the *Gammacoronavirus* genera. In solving these structures, we identified conserved RTC replication cofactor functions and requirements, while also demonstrating the potential for genera specific pathways of RTC assembly. This work demonstrates the importance of studying a broad range of CoVs.

CoVs unique ability to proofread has been well documented but the mechanism by which it occurs remains unknown. Here I present the substrate requirements for the

interaction of the CoV RTC and proofreading complex. Further, I have narrowed down the potential interaction site of these two complexes. This work provides critical insight into the unique mechanism of CoV proofreading.

To aid in the development of CoV antivirals we solved the structure of a SARS-CoV-2 RTC that's elongation is stalled by an arabinose nucleotide. In solving this structure, we identified that arabinose nucleotides are potent inhibitors of the CoV RTC. To our knowledge the use of arabinose nucleotides as CoV antivirals has not been previously tested. Our work demonstrates that these nucleoside analogues have the potential to be used as templates for antiviral drug development.

The work presented here provides critical insight into the mechanisms of CoV replication and proofreading, aiding in our ability to design more potent and broadly acting CoV antivirals, helping treat current and future CoV induced diseases.

Chapter 1: The molecular mechanisms of coronavirus replication

Authors: Thomas K. Anderson¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A. wrote this chapter with input and editing by R.N.K.

Coronaviruses as human and animal pathogens

Coronaviruses (CoVs) are a subfamily of viruses in the viral family *Coronaviridae* within the order *Nidovirales* (**Figure 1**) (1). The CoV subfamily is composed of four distinct genera: *Alpha*-, *Beta*-, *Gamma*-, and *Delta*-CoVs (**Figure 1**) (2). In humans, CoVs infect the respiratory tract and can cause a wide range of disease severities (2). There are four common cold coronaviruses that are endemic human pathogens, two from the *AlphaCoV* genera (HCoV-229E and NL63) and two from the *BetaCoV* genera (OC43 and HKU1) (**Figure 1**) (2,3). Since the turn of the century three *BetaCoVs* have emerged and caused outbreaks of severe respiratory disease in humans. SARS-CoV emerged in 2002 and infection could lead to symptoms of severe respiratory disease and pneumonia (4,5). The SARS-CoV outbreak resulted in >8,000 cases and >800 deaths (4-6). MERS-CoV emerged in 2012 and still causes infections through sporadic outbreaks mostly confined to the Middle East (7-9). Infection with MERS-CoV is associated with acute respiratory distress, pneumonia, possible kidney failure, and high mortality rates averaging around 35% (9,10). SARS-CoV-2 spilled over from animal reservoirs in 2019 and is the causative agent of the COVID-19 pandemic which has led to over 100 million cases and 1 million deaths in the U.S. alone (11,12). Symptoms associated with COVID-19 include atypical pneumonia, severe respiratory distress, and possible respiratory failure (13). The onset of the COVID-19 pandemic led to the rapid development and implementation of vaccines against SARS-CoV-2. While these new vaccines have proven effective at reducing SARS-CoV-2 related hospitalizations and deaths, the continued evolution of viral variants is a growing concern for vaccine effectiveness (14,15). In addition to the vaccines, three antiviral drugs have been approved for the treatment of COVID-19: Remdesivir,

Molnupiravir and Paxlovid (16). These antivirals will be described in greater detail in the **Coronavirus antivirals** section of this chapter. Outside of the three emergent *BetaCoVs* described above, a feline-canine recombinant *AlphaCoV*, canine-CoV human-pneumonia 2018 (CCoV-HuPn-2018), was isolated from human patients with pneumonia in the U.S. and Malaysia (17-19). More recently, the first reports of *DeltaCoV* infections in humans were reported as human adapted strains of porcine delta CoV (PDCoV) were isolated from children in Haiti (20). As CoVs from several genera continue to pose a threat to societal health we should continue to focus our research efforts into understanding CoV replication to aid in treating current and future CoV induced diseases (**Figure 1**).

In addition to being human pathogens, CoVs are important animal pathogens that can have detrimental impacts on the agricultural industry. The first CoV to be discovered was the *GammaCoV* infectious bronchitis virus (IBV) which is a prominent threat to chickens and consequently the chicken and egg production industries (**Figure 1**) (21,22). IBV primarily infects the respiratory tract, but some strains can spread to and infect the reproductive tract and kidneys (23). Infection of the kidneys can cause kidney lesions and nephritis (21,23,24). Infection of the female reproductive tract can lead to reduced egg quality which can be detrimental to the egg industry (21,23,24). Although vaccines have been developed for IBV, the broad genetic diversity of IBV strains and their rapid evolution have rendered vaccines ineffective at preventing IBV infection (25,26). Because of this IBV remains a persistent threat to the chicken and egg industries.

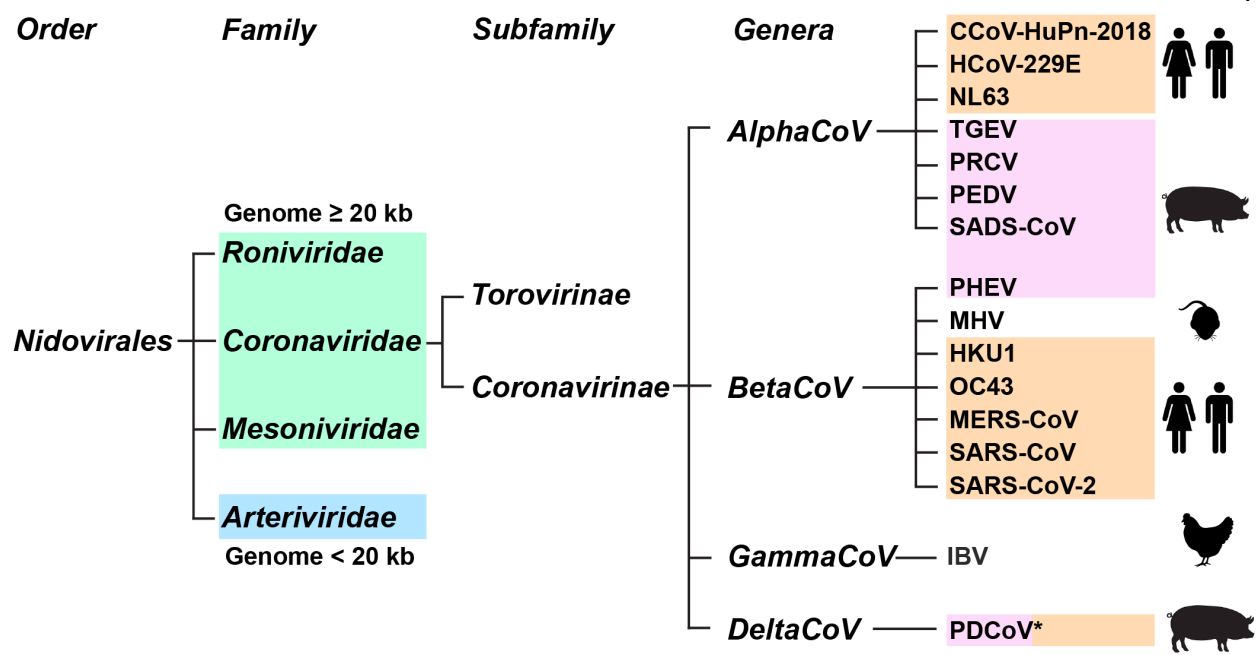


Figure 1, coronavirus taxonomy: Coronaviruses are a subfamily of viruses in the viral family *Coronaviridae* within the viral order *Nidovirales*. Two other families within *Nidovirales* have large genomes ≥20 kb (green) while one family, *Arteriviridae* (blue), have genomes <20 kb. Coronavirus species are split into four genera and infect numerous different hosts including humans (orange) and pigs (pink).

Beyond IBV there are several more CoVs that impact the agricultural industry, including several pathogenic pig CoVs (**Figure 1**). Currently there are six known pig CoVs, this includes four *AlphaCoVs* (porcine epidemic diarrhea virus (PEDV), transmissible gastroenteritis virus (TGEV), severe acute diarrhea syndrome CoV (SADS-CoV), and porcine respiratory CoV (PRCV)), one *BetaCoV* (porcine hemagglutinating encephalomyelitis virus (PHEV)) and one *DeltaCoV* (porcine deltacoronavirus (PDCoV)) (**Figure 1**) (27,28). PEDV was first identified in Europe in the 1970s but later spread to Asia in 1983, and the U.S. in 2013 (29-34). Following its emergence in the U.S. in 2013 a highly pathogenic strain of PEDV quickly became one of the most deadly and important global pig pathogens (28,35,36). PEDV is commonly spread by a fecal-oral route but can be transmitted from infected sows to piglets during suckling (37). PEDV infects the intestinal track and causes symptoms of vomiting, diarrhea, and dehydration (37). PEDV is highly contagious and outbreaks on a pig farm will result in almost all pigs being infected (37,38). While PEDV infection in adult pigs has low mortality rates, piglets' mortality rates can be as high as 100%, which is a serious threat to the swine industry (37,38). Like IBV, PEDV rapidly evolves, making vaccine design difficult and often ineffective (39). Another difficulty in PEDV vaccine design is that piglets are most susceptible to PEDV before they are weaned. At this young age, piglets cannot be effectively vaccinated and acquire their antibodies during suckling, further complicating vaccine implementation (39). IBV and PEDV are just two examples of the detrimental impacts that CoVs can have on the agricultural industry and exemplify the need to better understand all CoVs, not just those that infect humans.

Coronavirus genome and replication

Coronaviruses have (+) single-stranded RNA (ssRNA) genomes with 5' caps and 3' polyadenylated tails (2,40). In the realm of RNA virology, CoVs have uniquely large ~30 kb genomes. Most RNA viruses with genomes of similar or larger size are related viruses in the *Nidovirales* order, including a planarian nidovirus with a genome >40 kb (**Figure 1**) (1,41). RNA virus genomes are considered small with the majority having genomes ≤ 20 kb. RNA viruses' are believed to limit their genome size because they exist near an error threshold, a theoretical limit to the number of mutations a virus population can withstand to successfully replicate (42-45). When a virus surpasses this threshold, they are at risk of error catastrophe, a state in which the virus is unable to maintain a population of functioning genomes, resulting in failed replication and the extinction of that virus population (42,46). RNA viruses exist at this threshold because of the inherent low fidelity of the RNA-dependent RNA polymerases (RdRPs) these viruses use to replicate their RNA genomes (45,47). The low fidelity of viral RdRPs is believed to be a tradeoff for fast nucleotide addition rates that allow RNA viruses to produce large quantities of genome copies during an infection (48). These large populations of closely related but genetically distinct virus genomes produced during an infection are referred to as a quasispecies (44,49,50). Existing as a quasispecies can be beneficial as it allows the virus to sample evolutionary space to promote viral replication while avoiding host immune responses (44). However, existing as a quasispecies near the edge of error catastrophe can be coopted to treat RNA virus infections with antivirals that increase mutation rates to lethality for the virus population (46,51). CoVs, and other nidoviruses, exist as anomalies by replicating their large ~30 kb RNA genomes with one of the fastest known RdRPs (1,52).

How CoVs were able to maintain their large RNA genomes while avoiding error catastrophe remained a mystery until it was discovered that CoVs have evolved unique mechanisms to evade error catastrophe, including using virally encoded proofreading machinery, the details of which will be described below (1). Interestingly, the recent discovery of flavivirus-like viruses with genomes >20 kb potentially disrupts the theory of the error-threshold (53,54). While a homologue to nsp14 has not been identified in these viruses, more work is needed to determine whether they encode their own error-correction machinery.

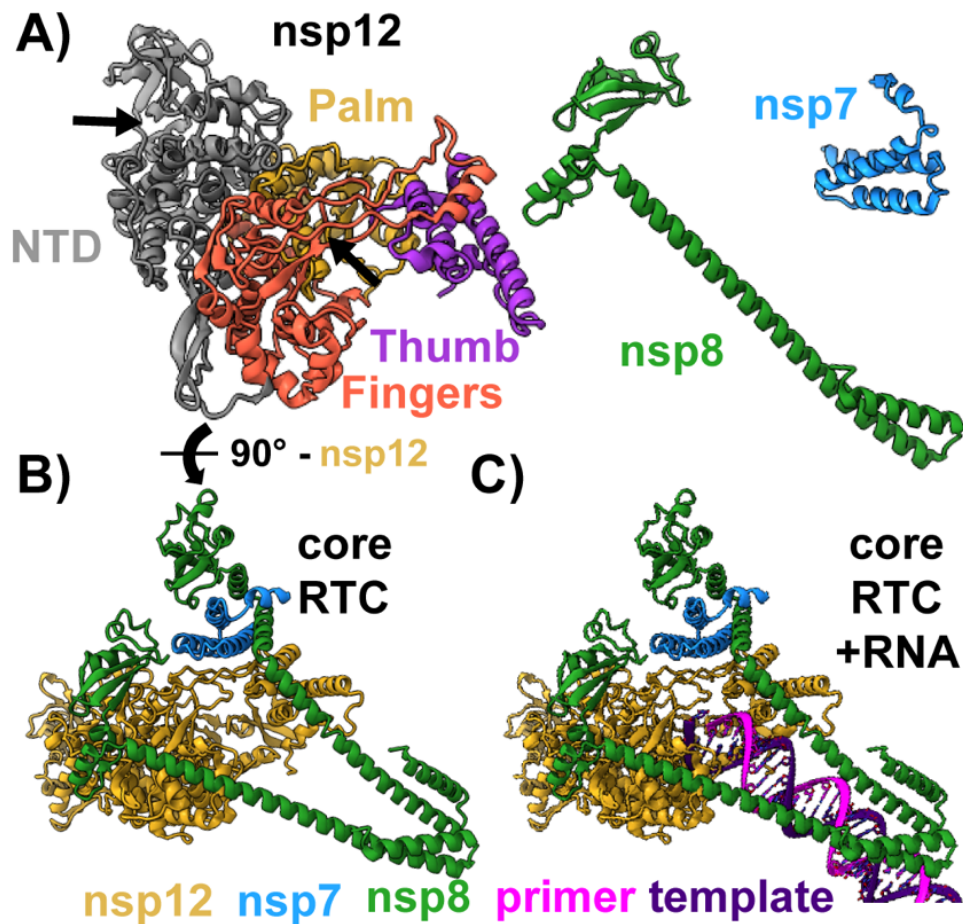


Figure 2, the CoV core-RTC: **A)** Cartoon representations of the viral cofactors nsp7 (blue) and nsp8 (green) and the viral polymerase nsp12, which is colored corresponding to the “cupped right hand” fingers, palm, and thumb domains. Nsp12s’ active sites are denoted with black arrows, with the NiRAN in the N-terminal domain (NTD) and RdRP in the palm. **B)** The core-RTC is composed of one nsp12, one nsp8, and one nsp7+nsp8 heterodimer. **C)** The N-terminal helical extensions of each nsp8 bind dsRNA leaving nsp12’s active site. Figures were made using PDB: 7KRP.

The 5' two-thirds of the CoV genome encodes two large open reading frames (ORFs), ORF1A and ORF1B, that encode two large polyproteins, pp1A and pp1AB (55). A -1 ribosomal frameshift at the end of ORF1A introduces the host ribosome into ORF1B allowing translation of the second and larger pp1AB (56,57). The two polyproteins are cleaved by two or three viral proteases into 11 or 15 viral non-structural proteins (nsps) (**Table 1**) (55). Once translated, the viral nsps are responsible for both modulating the host cell to support viral replication and producing viral RNA for translation and virion assembly (55,58). Three of the nsps remodel the host endoplasmic reticulum membrane to form double membrane vesicles (DMVs) which are the sites of virus replication (**Table 1**) (59-62). The viral protein nsp3 has been shown to form a transmembrane pore that spans the membranes of DMVs (63). This pore is believed to allow the egress of viral messenger RNA (mRNA) and genomic RNA (gRNA) from DMVs (63). Within DMVs, several more nsps assemble into the viral replication-transcription complex (RTC) that is responsible for the production and modification of viral RNA products (62,64-66). During replication, CoVs use the unique mechanism of discontinuous strand synthesis (DSS) to synthesize sub-genomic RNAs (sgRNAs) that are used for the translation of viral accessory and structural proteins (67). DSS occurs during (-) RNA synthesis when the RTC template switches from transcriptional regulatory sequences (TRS) separating each of the structural and accessory genes to a matching TRS at the 5' end of the (+) genome (67,68). This process creates a set of nested (-) sgRNAs that are used as templates for (+) sgRNA synthesis. Each (+) sgRNA has a unique 5' terminal gene that allows for translation of that structural or accessory protein.

The central component of the CoV RTC is the viral RdRP nsp12 (**Figure 2A**, **Table 1**) (58,69). In addition to its RdRP activity, nsp12 has a second active site located within its N-terminal domain (NTD) termed the nidovirus RdRP-associated nucleotidyl transferase (NiRAN) (**Figure 2A**) (70). The NiRAN domain is a unique nucleotidyl transferase that is involved in viral mRNA capping and is conserved across the *Nidovirales* order (71,72). It was first shown for SARS-CoV that nsp12 requires two viral cofactors, nsp7 and nsp8, for optimal RNA synthesis activity *in vitro* (**Table 1**) (73). This result has been reproduced for SARS-CoV-2 and the *Alphacoronavirus* PEDV (74,75). The first structure of a CoV RTC was of SARS-CoV nsp12 with cofactors nsp7 and nsp8 (76). This structure revealed a surprising complex stoichiometry of one nsp12 bound to one nsp8 protomer and one nsp7 + nsp8 heterodimer (**Figure 2B**) (76). To distinguish the nsp8s they are referred to as nsp8_F (protomer) and nsp8_T (nsp7 heterodimer) based off whether they bind the fingers (F) or thumb (T) domain of the polymerase. This complex of nsp12 with nsp7 and two nsp8s is considered the CoV core-RTC as it contains the minimum components needed to make a complex capable of processive RNA synthesis (**Figure 2**) (52,73). The CoV core-RTC is the fastest known RdRP with nucleotide addition rates up to ~170 nt/sec (52). This replication speed is interesting as RNA viruses are known to sacrifice replication fidelity for replication rates (48). Recent work on the SARS-CoV-2 core-RTC demonstrated that CoVs can sustain such fast replication without greatly reducing fidelity due to two conserved mutations in the polymerase active site (74).

Since their initial discovery as replication cofactors, the dynamics and structures of nsp7 and nsp8 have been further elucidated both *in vitro* and *in vivo*. The cofactor nsp8 has been shown to form homo-oligomers that can be de-oligomerized by nsp7 (77-80).

These nsp7+nsp8 complexes are then able to form higher order oligomers (77-80). The types of oligomers sampled by nsp8 and nsp7+nsp8 complexes vary depending on the CoV genera (81). Work that identified nsp7 and nsp8 as cofactors for nsp12 used a nsp7-nsp8 fusion protein to maximize RdRP activity *in vitro* (73). This fusion protein was constructed with the C terminus of nsp7 fused to the N terminus of nsp8 by 6 histidine residues. The functionality of this construct validated prior reverse genetics experiments that demonstrated that a mutant *BetaCoV* with a protease resistant nsp7-nsp8 cleavage site could replicate in tissue culture (82). Mutational screens of nsp8 to test the ability of mutants to bind nsp12 and promote RNA synthesis identified the interesting mutation of nsp8-K58A (73). This mutant nsp8 could bind nsp12 with similar affinity to wildtype but was incapable of promoting RNA synthesis (73). Later structures of the SARS-CoV-2 core-RTC revealed that this nsp8 residue binds directly to dsRNA exiting the polymerase active site, promoting RTC assembly with RNA (**Figure 2**) (83,84). Further structures of the SARS-CoV and SARS-CoV-2 RTCs revealed that the C terminus of nsp7 was positioned far away from the N terminus of nsp8, making the functionality of the nsp7-nsp8 fusion perplexing (**Figure 2**) (83,84). With this insight, a different fusion protein with a flipped sequence (nsp8-nsp7 instead of nsp7-nsp8) was developed. This new fusion protein allowed for a more realistic core-RTC to be assembled with the nsp8-nsp7 heterodimer binding the thumb domain, and free nsp8 binding the fingers domain (74). This nsp8-nsp7 fusion protein was used to delineate the two nsp8s' contributions to complex assembly, revealing that the binding of nsp8_F to nsp12 is the rate limiting step in core-RTC formation (74).

Protein	Enzyme activity and/or protein function	Reference(s)
nsp1	Host translation inhibition	(85,86)
nsp2	Innate immune regulation	(87-89)
nsp3	ER remodeling, DMV formation, DMV pore, papain like protease	(63,90,91)
nsp4	ER remodeling, DMV formation	(60,91)
nsp5	Main protease, polyprotein cleavage	(92,93)
nsp6	ER remodeling, DMV formation	(59,94)
nsp7	Replication cofactor for nsp12	(73,75)
nsp8	Replication cofactor for nsp12	(73,84)
nsp9	RNA capping cofactor	(71,95)
nsp10	Cofactor for nsp14 and nsp16	(96)
nsp11	Polyprotein cleavage byproduct	(97)
nsp12	RNA-dependent RNA Polymerase, NiRAN	(69,73)
nsp13	5'-3' RNA helicase, RNA 5' triphosphatase	(98,99)
nsp14	3'-5' exoribonuclease, N7-methyltransferase	(100,101)
nsp15	Endoribonuclease (uridine specific)	(102)
nsp16	2'O-methyltransferase	(103)

Table 1, the CoV nsps: Listed are the 16 nsps and their roles in CoV replication.

As mentioned above CoVs encode over a dozen nsps that are involved in genome replication (**Table 1**). Among these, CoVs encode the enzymes necessary to assemble 5' cap 0, 1 and 2 structures for efficient translation of viral mRNAs (**Figure 3, Table 1**). This includes nsp13, a 5' triphosphatase that dephosphorylates the 5' end of viral RNAs to prepare them for the addition of the inverted 5'-5' linked GTP-cap via the NiRAN domain of nsp12 (**Figure 3, Table 1**) (58). The viral cofactor nsp9 is known to bind the NiRAN and is implicated in RNA capping (71). The RNA cap can then be methylated by the N7-methyltransferase nsp14 to form a cap 0 structure, and the first two nucleotides of mRNA can be 2'OH methylated by nsp16 to form cap 1 and cap 2 structures (**Figure 3, Table 1**) (58). In addition to its role in capping, nsp13 is a 5'-3' RNA helicase that has been shown to be critical for CoV replication (**Table 1**) (104,105). CoVs encode an endonuclease, nsp15, that cleaves dsRNA 3' of uridines to limit innate immune stimulation (**Table 1**) (106,107). A second CoV ribonuclease is encoded in the bifunctional enzyme nsp14, that contains both a 3'-5' exoribonuclease domain and the aforementioned N7-methyltransferase (**Table 1**) (108). Through its exonuclease activity, nsp14 mediates CoV proofreading (109,110). CoVs are unique among RNA viruses in their ability to proofread, which also makes antiviral nucleoside analogues (NAs) such as Ribavirin and 5-fluorouracil ineffective against CoVs (110,111). The cofactor nsp10 promotes the enzymatic activities of nsp14 and nsp16, and it has been shown that a complex of nsp10+nsp14+nsp16 could exist to mediate cap methylations (**Table 1**) (96,112). These viral nsps, and more, are believed to form large multimeric enzymatic RTCs within viral DMVs, but how they interact, and function together remains largely unknown.

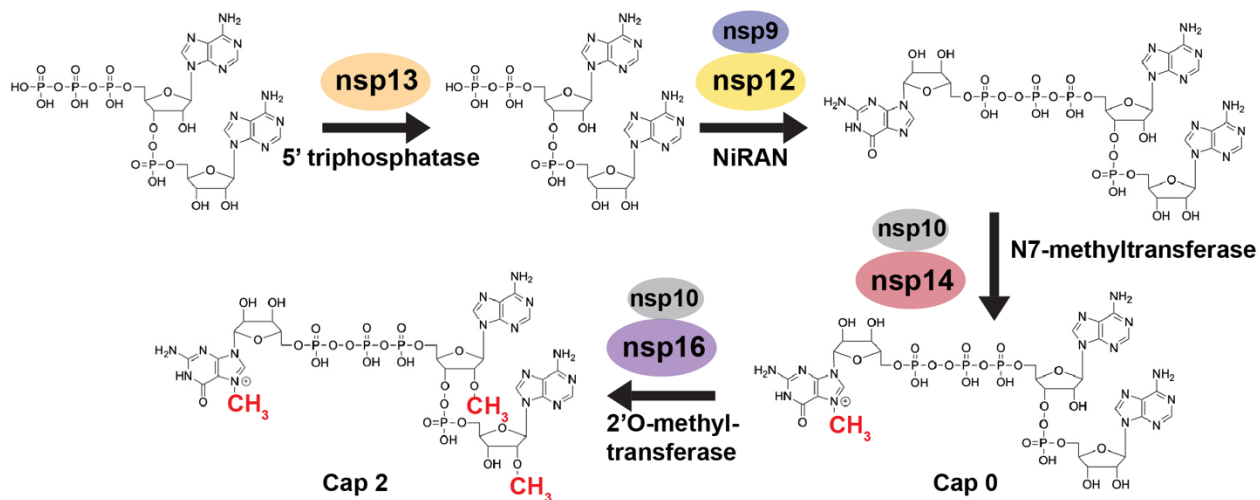


Figure 3, CoV capping mechanism: CoVs encode all enzymes necessary to add cap 0, cap 1, and cap 2 structures on viral mRNAs. Starting from the top left and moving clockwise: nsp13 contains a 5' triphosphatase that can remove the gamma phosphate from 5'-triphosphorylated RNAs. The NiRAN domain can then mediate the addition of the 5'-5' linked guanosine cap. This is believed to be catalyzed by nsp12's NiRAN domain, using the cofactor nsp9, the full mechanism of this has yet to be fully elucidated. The 5' guanosine N7 can be methylated by nsp14 to form a cap 0 structure. The first two mRNA nucleotide's 2'OHs can be methylated by nsp16 to form cap 1 (first nucleotide methylated) and cap 2 (first and second nucleotide methylated). Nsp10 is a cofactor for both nsp14 and nsp16.

Structural insights into coronavirus replication

Structures of the SARS-CoV(-2) core-RTC have revealed that nsp12 has the canonical (+) RNA virus RdRP shape that is best described as a cupped right hand (**Figure 2A**) (76,84,113,114). Following this metaphor, the polymerase is split into three domains: the fingers, palm, and thumb with the RdRP active site located within the palm (**Figure 2A**) (114). The three domains are composed of seven conserved motifs (A-G) (113). Akin to other (+) ssRNA virus RdRPs the “fingertips” of nsp12’s fingers domain reach across the polymerase “palm” to contact the “thumb” and enclose the RdRP active site (**Figure 2**) (115). This interaction creates two channels in the RdRP: one for RNA to bind and translocate through and a second for NTP entry (114,115). Within RdRP active sites the +1 templating RNA is base stacked with the template base directly upstream of it, positioning the template base for NTP binding within the polymerase active site (83,84,116). Positioning of the +1-templating base in the active site differs from other polymerases (such as DNA polymerases) that use a secondary NTP binding site (117). After NTP binding the RdRP active site closes, most notably by a shift in motif A (115,116). This closure orients active site residues and two Mg^{2+} cofactors to coordinate the nucleophilic attack of the NTP’s alpha-phosphate by the nascent RNA 3’OH (115,118). After NMP addition, motif A shifts back to open the active site and allow the release of the pyrophosphate product (118). RNA is then translocated through the polymerase via a Brownian ratchet mechanism, positioning the next downstream template base for NTP binding (118). In addition to detailing the mechanisms of NTP binding, structures of the SARS-CoV-2 core-RTC bound to RNA revealed that each of the RTCs’ nsp8s have long helical extensions that bind dsRNA exiting the polymerase active site (**Figure 2**) (83,84).

These RNA binding extensions were predicted to promote processive replication by the RTC (84). More recent work on SARS-CoV-2 and PEDV have demonstrated that only one nsp8 RNA binding domain is needed for RNA synthesis *in vitro* (74,75).

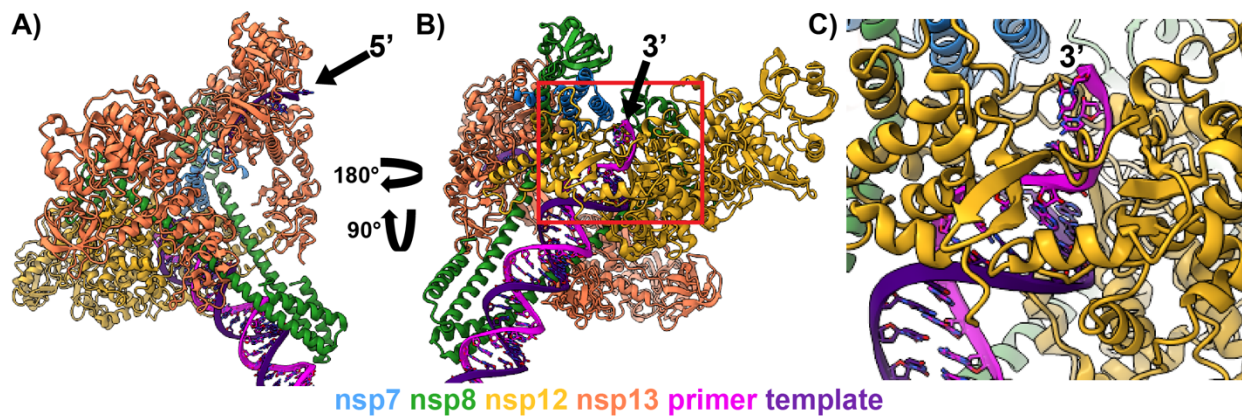


Figure 4, nsp13 mediated backtracking: **A)** Two viral helicases, nsp13 (coral), can associate with the core-RTC, with each being scaffolded to the complex by a different nsp8. The nsp13 bound to nsp8_T binds the 5' end of downstream template RNA (purple) that would be entering the RdRP active to be copied during replication. **B)** The CoV RTC can bind a forked RNA substrate with mismatches at the 3' end of primer. This results in the 3' mismatches being backtracked through nsp12's NTP channel. **C)** Zoomed in view of the 3' end of nascent strand RNA being extruded out the NTP channel. The model used for figure making was PDB: 7KRO.

Structures of the SARS-CoV-2 RTC with additional nsps beyond those of the core-RTC (nsp7, nsp8, and nsp12) have begun to shine a light on RTC dynamics and function. The cofactor nsp9 has been shown to bind the NiRAN domain of nsp12, with nsp9's N terminus inserting into the NiRAN active site (71). Further biochemical and structural experiments have begun to elucidate a mechanism of CoV RNA capping that involves the NiRAN NMP- or RNA-ylating nsp9 as a capping intermediate (95). The addition of the viral helicase, nsp13, to structures of the SARS-CoV-2 RTC had a surprising complex stoichiometry of one RTC bound to two helicases (**Figure 4A**) (119,120). Each of the nsp13s was scaffolded to the RTC by one of the nsp8 protomers (**Figure 4A**) (119,120). The nsp13 scaffolded by nsp8_T was bound to the 5' end of template RNA that would be entering the RdRP active site during elongation (**Figure 4A**) (120). This structure produced a conundrum as the direction of nsp13, 5'-3' on the template strand, opposes that of nsp12, 3'-5' on the template. This observation led to the prediction that nsp13 could be a molecular motor for RTC backtracking. To test this hypothesis a structure of the SARS-CoV-2 RTC bound to an RNA primer-template pair with 5 mismatched base pairs at the primer 3' was solved (**Figure 4**) (121). The 3' primer mismatches were shown to be extruded out of nsp12's NTP channel, homologous to mechanisms of backtracking by the eukaryotic RNA polymerase II (**Figure 4B and C**) (122). Although 5 consecutive mismatches are unlikely to form during replication, similar frayed RNA substrates are commonly used to isolate polymerases in backtracked states for structural characterization (123,124). Further biochemical testing confirmed that nsp13 could induce RTC backtracking, pushing the 3' end of nascent RNA through the NTP channel (121). Currently, the purpose, signal, and regulation of backtracking remain unknown. It

has been proposed that backtracking could play an important role in CoV discontinuous strand synthesis or proofreading. Backtracking to an exonuclease active site (like that of nsp14) after mismatch introduction is a mechanism used by other eukaryotic and prokaryotic polymerases (122,125,126). In an attempt to elucidate the mechanism of CoV proofreading, a structure of a SARS-CoV-2 RTC with nsp13, nsp9, nsp10, and nsp14 bound to the core-RTC was solved (127). Due to weak interactions between the RTC and nsp10+nsp14, a nsp9-nsp10 fusion protein was used to stably assemble the complex. This fusion resulted in the dimerization of these large RTCs (127). An important caveat to this structure is that the main interaction between the proofreading complex (nsp10+14) and the core-RTC is mediated by the nsp9-nsp10 fusion, indicating that complex assembly might be an artifact of the fusion protein. While interesting, further structural and biochemical characterization is needed to determine the validity of this model. Using the rapid increase of CoV RTC structural data, a molecular modeling and protein docking study proposed an RTC superstructure composed of an nsp15 (viral endonuclease) hexameric core that scaffolds two nsp10+nsp14+nsp16 trimers and six RTCs (nsp7, nsp8, nsp12...) (128). In the superstructure viral RNA is predicted to thread across this complex into different enzymatic active sites. While the validity of this model has not been tested, it highlights the potential complexity and scale of CoV RTCs that has yet to be fully explored and determined.

All the structures described above are of either SARS-CoV or SARS-CoV-2, two very closely related *BetaCoVs*. Further, *in vitro* characterization of non-betaCoV polymerases is very limited with few reports of robust polymerase activity outside of

BetaCoVs. By limiting our studies to two closely related *BetaCoVs* we are unintentionally limiting our ability to fully comprehend CoV replication.

Coronavirus proofreading

As described in the **Coronavirus genome and replication** section CoVs have uncommonly large RNA genomes. The ability of CoVs to replicate these large genomes without entering error catastrophe puzzled virologists until a 3'-5' exoribonuclease was identified in the CoV protein nsp14 (101). Nsp14 is classified as a DE(D/E)Dh exonuclease, which is the same superfamily of exonucleases as other proofreading enzymes such as the Klenow fragment of DNA Polymerase I (101,129). *In vitro* characterization of nsp14 has confirmed its 3'-5' exoribonuclease activity, and that it requires the cofactor nsp10 for optimal activity (**Table 1**) (101,130,131) (111). A recombinant SARS-CoV nsp10+nsp14 complex was shown to degrade mismatched 3' nucleotides from RNA, representative of a frayed RNA substrate after misincorporation (132). Nsp14 can interact with the core-RTC (nsp7+nsp8+nsp12) in *in vitro* pulldown assays, but these experiments excluded nsp10 from reactions limiting our ability to know nsp10's role in RTC assembly (73). Several studies have knocked out nsp14's exonuclease activity in different CoVs and produced varying results. Knockout of exonuclease activity in the *BetaCoVs* SARS-CoV-2 and MERS-CoV produced non-viable viruses (133). In the case of the *AlphaCoV* 229E, knockout resulted in greatly reduced RNA synthesis and no progeny virus production (101). In contrast, knockout of exonuclease activity in the *BetaCoVs* MHV and SARS-CoV produced viable viruses with increased mutation rates up to 15x that of wildtype virus (100,134). While these results vary between viruses, they demonstrate the importance of nsp14 and its role in

proofreading for CoV replication. Nsp14 proofreading also renders commonly used nucleoside analogues such as Ribavirin and 5-Fluorouracil ineffective at blocking CoV replication (110,111). Mutations that knockout nsp14's exonuclease or block nsp10's binding to nsp14, render CoVs sensitive to 5-fluorouracil treatment (110,135). While it is known that nsp14 is responsible for proofreading during replication, the mechanism by which proofreading occurs is unknown.

There are several theories on the mechanism of CoV proofreading. One theory predicts that after misincorporation the polymerase disassociates from the RNA, providing a frayed RNA substrate for nsp10+nsp14 association and excision (**Figure 5**). Another possible mechanism for CoV proofreading involves RTC backtracking mediated by nsp13 (**Figure 5**). The structures of the backtracked SARS-CoV-2 RTC with nsp13, placed the 3' end of nascent RNA out the polymerase NTP channel (**Figure 4 B and C**) (121). In the case of a misincorporation, a newly incorporated mismatch could stimulate backtracking via nsp13, presenting the mismatch nucleotide out the NTP channel for excision by nsp14 (**Figure 5**). This mechanism of proofreading would allow for the controlled degradation of the mismatch as the remaining portion of nascent RNA would remain bound to nsp12. Similar mechanisms of proofreading that involve the movement of the newly incorporated mismatch to a 3'-5' exonuclease active site are used by other polymerases such as eukaryotic and prokaryotic DNA polymerases (125,126,136).

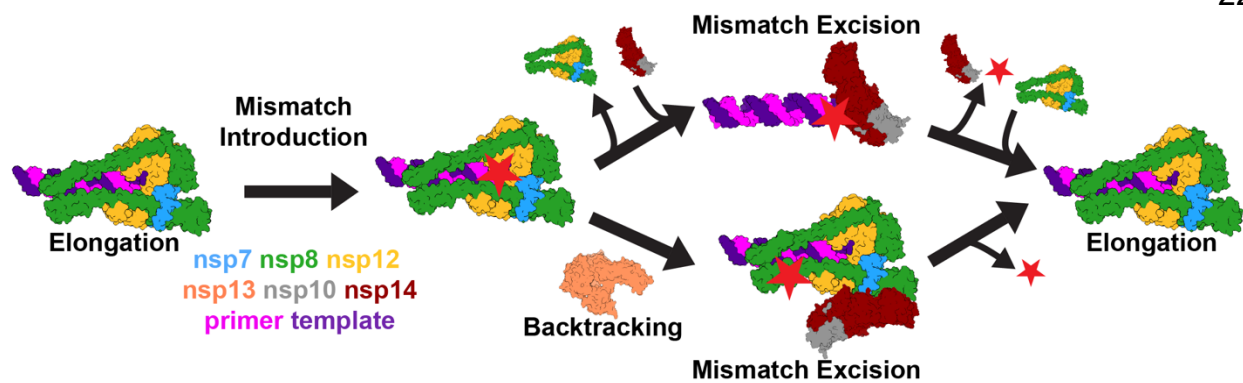


Figure 5, theoretical models of CoV proofreading: Illustrated are two hypotheses as to the mechanism(s) by which CoVs proofread. After a mismatch (red star) is introduced by nsp12 the RTC could either disassemble and/or disassociate from the RNA (top path) allowing nsp10+nsp14 to associate and excise the mismatch. Another possible mechanism involves backtracking mediated by nsp13 after mismatch introduction (bottom path). After backtracking, the mismatch at the 3' end of the nascent strand would be presented out of the NTP channel allowing nsp10+nsp14 to excise the mismatch. For simplicity, nsp13 and nsp10+nsp14 were not included on elongating RTCs as it is not known whether they associate with the RTC during elongation or just during backtracking or proofreading.

Coronavirus antivirals

The onset of the COVID-19 pandemic led to the rapid development of several highly effective SARS-CoV-2 vaccines. While these vaccines have played an important role in reducing the impact of the pandemic, antiviral drugs are still required to treat cases of COVID-19 to limit virus spread and reduce disease severity. Since its onset, three drugs have been approved for the treatment of COVID-19: Remdesivir, Molnupiravir, and Paxlovid (**Figure 6**). The first two of these, Remdesivir and Molnupiravir, target the viral polymerase, nsp12, while Paxlovid targets the viral main protease, nsp5 (**Figure 6**). Early in the pandemic, Remdesivir showed promise as a nucleoside analogue against SARS-CoV-2 in tissue culture experiments (137). One such study reported IC₅₀ values lower than 10 nM in primary human airway cells (137). The first clinical trial of Remdesivir in the treatment of COVID-19 presented a 33% decrease in recovery time from 15 to 10 days, and because of this, Remdesivir received emergency use authorization (138,139). Further studies revealed that Remdesivir did not significantly reduce mortality rates when compared to the standard of care and the WHO later recommended against its use (138). Molnupiravir, a hyper-mutagenic cytosine analogue, was shown to be a potent antiviral against several CoVs (including SARS-CoV-2) in tissue culture models (**Figure 6**) (140,141). In ferret transmission studies, Molnupiravir reduced SARS-CoV-2 replication and transmission between ferrets (142). A clinical trial of Molnupiravir intervention demonstrated a reduced risk of COVID-19 related hospitalizations by ~50% in unvaccinated adults. Molnupiravir received emergency use authorization in 2021 (143,144). While Molnupiravir remains approved for the treatment of COVID-19, its efficacy is still debated because of its inability to reduce hospitalizations in vaccinated,

high-risk adults (145). The protease inhibitor Paxlovid is composed of two drugs: the first being the CoV main protease inhibitor Nirmatrelvir, and the second, Ritonavir, decreases the metabolic processing of Nirmatrelvir to improve its efficacy. Nirmatrelvir was first identified to be an effective SARS-CoV-2 antiviral in tissue culture and mouse model experiments (146). Clinical trials of Paxlovid in the treatment of COVID-19 have demonstrated reduced severe COVID-19 related hospitalizations or deaths by up to ~90%, solidifying it as the most effective SARS-CoV-2 antiviral yet (147).

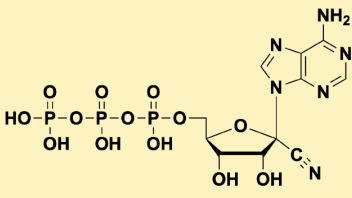
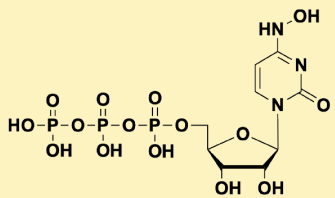
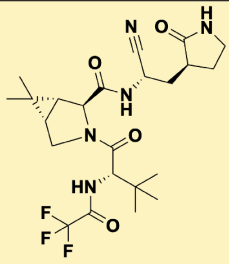
Drug	Remdesivir	Molnupiravir	Nirmatrelvir (Paxlovid)
Structure			
Target	nsp12 RdRP	nsp12 RdRP	nsp5 protease

Figure 6, CoV antivirals: The structures and viral targets of the three antivirals to receive authorization for the treatment of COVID-19 since the onset of the pandemic are provided. The two targeting nsp12 (Remdesivir and Molnupiravir) are shown as their active tri-phosphorylated forms.

Structures of the SARS-CoV-2 core-RTC bound to nucleoside analogues have identified the drugs' mechanisms of action in blocking virus replication. A series of structures of Remdesivir, an adenosine mimetic with a 1' cyano substitution, documented that active site residue nsp12-Ser861 sterically clashes with the 1' cyano, blocking Remdesivir translocation in the nascent strand from the +3 to +4 position during RNA elongation, stalling the RTC (148). The antiviral Molnupiravir was shown to base-pair with adenosine and guanosine in nsp12's active site, providing structural insight into its mechanism of hyper-mutagenesis (149).

Nucleoside analogue antivirals are commonly used and effective antiviral treatments against numerous types of viruses. For example, Acyclovir and Sofosbuvir are two highly effective NAs used to treat herpes simplex virus (DNA virus) or hepatitis C virus (RNA virus) infections, respectively. These are just two examples from a multitude of NAs that have been developed against DNA and RNA viruses over the years. CoVs are unusually difficult to design effective NAs against due to their proofreading capability. This is best demonstrated by the two antivirals that have been ineffective in the clinic, Remdesivir and Molnupiravir, which are nucleoside analogues. To be effective against CoVs a NA must both be efficiently incorporated by the polymerase nsp12 while avoiding/blocking proofreading by nsp14. While Paxlovid has proven effective in treating COVID-19, more effective antivirals, including nucleoside analogues, will be required for the treatment of CoV induced diseases.

Concluding remarks

By encoding over a dozen enzymes and cofactors to replicate and modify their RNA genomes, coronaviruses are remarkably complicated and interesting RNA viruses.

Over the last couple of decades, substantial work has been done to identify and characterize these viral enzymes and the roles they play in replication. Increased interest spurred by the COVID-19 pandemic has resulted in an unprecedented burst of research into the field of corona-virology. While our understanding of CoV replication grows, we continue to realize how little we truly know about these viruses. Among the mechanisms that are not yet elucidated are proofreading and discontinuous strand synthesis, two processes unique to CoV replication. While the evidence of CoV backtracking provides a possible foundation for proofreading or DSS, the true role of RTC backtracking remains unknown. Beyond the RTC, our understanding of how all the nsps within an infected cell interact to mediate replication within the context of viral double membrane vesicles is almost completely unknown. In addition, most of the research on CoVs has focused on SARS-CoV, SARS-CoV-2, MHV, and MERS-CoV, all of which are *BetaCoVs*. Excluding viruses from the *Alpha-*, *Gamma-*, and *Delta-CoV* genera from our research efforts is detrimental to gaining a complete understanding of CoV replication. The more we know about CoV replication, the more opportunities we will have for the development of broadly acting and highly effective antivirals, better preparing us for current and future CoV induced diseases.

References:

1. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res*, **117**, 17-37.
2. Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, **1282**, 1-23.
3. Monto, A.S., DeJonge, P.M., Callear, A.P., Bazzi, L.A., Capriola, S.B., Malosh, R.E., Martin, E.T. and Petrie, J.G. (2020) Coronavirus Occurrence and Transmission Over 8 Years in the HIVE Cohort of Households in Michigan. *J Infect Dis*, **222**, 9-16.

4. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.
5. Sims, A.C., Baric, R.S., Yount, B., Burkett, S.E., Collins, P.L. and Pickles, R.J. (2005) Severe acute respiratory syndrome coronavirus infection of human ciliated airway epithelia: role of ciliated cells in viral spread in the conducting airways of the lungs. *J Virol*, **79**, 15511-15524.
6. Davis, I.M. (2021) SARS-CoV: Lessons learned; opportunities missed for SARS-CoV-2. *Rev Med Virol*, **31**, 1-6.
7. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. and Fouchier, R.A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, **367**, 1814-1820.
8. de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R.A., Galiano, M., Gorbalenya, A.E., Memish, Z.A. *et al.* (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, **87**, 7790-7792.
9. Widagdo, W., Okba, N.M.A., Stalin Raj, V. and Haagmans, B.L. (2017) MERS-coronavirus: From discovery to intervention. *One Health*, **3**, 11-16.
10. Chafekar, A. and Fielding, B.C. (2018) MERS-CoV: Understanding the Latest Human Coronavirus Threat. *Viruses*, **10**.
11. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265-269.
12. (2024) WHO COVID-19 DASHBOARD.
13. Hu, B., Guo, H., Zhou, P. and Shi, Z.L. (2021) Characteristics of SARS-CoV-2 and COVID-19. *Nat Rev Microbiol*, **19**, 141-154.
14. Lyke, K.E., Atmar, R.L., Islas, C.D., Posavad, C.M., Szydlo, D., Paul Chourdury, R., Deming, M.E., Eaton, A., Jackson, L.A., Branche, A.R. *et al.* (2022) Rapid decline in vaccine-boosted neutralizing antibodies against SARS-CoV-2 Omicron variant. *Cell Rep Med*, **3**, 100679.
15. Cohen, J. (2021) Omicron sparks a vaccine strategy debate. *Science*, **374**, 1544-1545.
16. Vangeel, L., Chiu, W., De Jonghe, S., Maes, P., Slechten, B., Raymenants, J., André, E., Leyssen, P., Neyts, J. and Jochmans, D. (2022) Remdesivir,

- Molnupiravir and Nirmatrelvir remain active against SARS-CoV-2 Omicron and other variants of concern. *Antiviral Res*, **198**, 105252.
17. Vlasova, A.N., Diaz, A., Dامتie, D., Xiu, L., Toh, T.H., Lee, J.S., Saif, L.J. and Gray, G.C. (2022) Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis*, **74**, 446-454.
 18. Silva, C.S., Mullis, L.B., Pereira, O., Jr., Saif, L.J., Vlasova, A., Zhang, X., Owens, R.J., Paulson, D., Taylor, D., Haynes, L.M. and Azevedo, M.P. (2014) Human Respiratory Coronaviruses Detected In Patients with Influenza-Like Illness in Arkansas, USA. *Virology*, **2014**, 1-8.
 19. Xiu, L., Binder, R.A., Alarja, N.A., Kochek, K., Coleman, K.K., Than, S.T., Bailey, E.S., Bui, V.N., Toh, T.H., Erdman, D.D. and Gray, G.C. (2020) A RT-PCR assay for the detection of coronaviruses from four genera. *J Clin Virol*, **128**, 104391.
 20. Lednicky, J.A., Tagliamonte, M.S., White, S.K., Elbadry, M.A., Alam, M.M., Stephenson, C.J., Bonny, T.S., Loeb, J.C., Telisma, T., Chavannes, S. *et al.* (2021) Independent infections of porcine deltacoronavirus among Haitian children. *Nature*, **600**, 133-137.
 21. Jackwood, M.W. (2012) Review of infectious bronchitis virus around the world. *Avian Dis*, **56**, 634-641.
 22. Hawn, S.a. (1931) First Report of Infectious Bronchitis in North Dakota, "An Apparently New Respiratory Disease of Chicks". *Journal of the American Veterinary Medical Association*, **78**, 413-422.
 23. Raj, G.D. and Jones, R.C. (1997) Infectious bronchitis virus: Immunopathogenesis of infection in the chicken. *Avian Pathol*, **26**, 677-706.
 24. Hoerr, F.J. (2021) The Pathology of Infectious Bronchitis. *Avian Dis*, **65**, 600-611.
 25. Bali, K., Bálint, Á., Farsang, A., Marton, S., Nagy, B., Kaszab, E., Belák, S., Palya, V. and Bányai, K. (2021) Recombination Events Shape the Genomic Evolution of Infectious Bronchitis Virus in Europe. *Viruses*, **13**.
 26. Tizard, I.R. (2020) Vaccination against coronaviruses in domestic animals. *Vaccine*, **38**, 5123-5130.
 27. Wang, Q., Vlasova, A.N., Kenney, S.P. and Saif, L.J. (2019) Emerging and re-emerging coronaviruses in pigs. *Curr Opin Virol*, **34**, 39-49.
 28. Lei, J., Miao, Y., Bi, W., Xiang, C., Li, W., Zhang, R., Li, Q. and Yang, Z. (2024) Porcine Epidemic Diarrhea Virus: Etiology, Epidemiology, Antigenicity, and Control Strategies in China. *Animals (Basel)*, **14**.

29. Wood, E.N. (1977) An apparently new syndrome of porcine epidemic diarrhoea. *Vet Rec*, **100**, 243-244.
30. Takahashi, K., Okada, K. and Ohshima, K. (1983) An outbreak of swine diarrhea of a new-type associated with coronavirus-like particles in Japan. *Nihon Juigaku Zasshi*, **45**, 829-832.
31. Chen, J.F., Sun, D.B., Wang, C.B., Shi, H.Y., Cui, X.C., Liu, S.W., Qiu, H.J. and Feng, L. (2008) Molecular characterization and phylogenetic analysis of membrane protein genes of porcine epidemic diarrhea virus isolates in China. *Virus Genes*, **36**, 355-364.
32. Chen, J., Wang, C., Shi, H., Qiu, H.J., Liu, S., Shi, D., Zhang, X. and Feng, L. (2011) Complete genome sequence of a Chinese virulent porcine epidemic diarrhea virus strain. *J Virol*, **85**, 11538-11539.
33. Cima, G. (2013) Fighting a deadly pig disease. Industry, veterinarians trying to contain PED virus, new to the US. *J Am Vet Med Assoc*, **243**, 469-470.
34. Jung, K. and Saif, L.J. Porcine epidemic diarrhea virus infection: Etiology, epidemiology, pathogenesis and immunoprophylaxis.
35. Chen, Q., Li, G., Stasko, J., Thomas, J.T., Stensland, W.R., Pillatzki, A.E., Gauger, P.C., Schwartz, K.J., Madson, D., Yoon, K.J. *et al.* (2014) Isolation and characterization of porcine epidemic diarrhea viruses associated with the 2013 disease outbreak among swine in the United States. *J Clin Microbiol*, **52**, 234-243.
36. Hanke, D., Pohlmann, A., Sauter-Louis, C., Höper, D., Stadler, J., Ritzmann, M., Steinrigl, A., Schwarz, B.A., Akimkin, V., Fux, R. *et al.* (2017) Porcine Epidemic Diarrhea in Europe: In-Detail Analyses of Disease Dynamics and Molecular Epidemiology. *Viruses*, **9**.
37. Zhang, Y., Chen, Y., Zhou, J., Wang, X., Ma, L., Li, J., Yang, L., Yuan, H., Pang, D. and Ouyang, H. (2022) Porcine Epidemic Diarrhea Virus: An Updated Overview of Virus Epidemiology, Virulence Variation Patterns and Virus-Host Interactions. *Viruses*, **14**.
38. Stevenson, G.W., Hoang, H., Schwartz, K.J., Burrough, E.R., Sun, D., Madson, D., Cooper, V.L., Pillatzki, A., Gauger, P., Schmitt, B.J. *et al.* (2013) Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J Vet Diagn Invest*, **25**, 649-654.
39. Gerdtts, V. and Zakhartchouk, A. (2017) Vaccines for porcine epidemic diarrhea virus and other swine coronaviruses. *Vet Microbiol*, **206**, 45-51.

40. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z. *et al.* (2020) Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, **183**, 730-738 e713.
41. Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A. and Gorbalenya, A.E. (2018) A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog*, **14**, e1007314.
42. Summers, J. and Litwin, S. (2006) Examining the theory of error catastrophe. *J Virol*, **80**, 20-26.
43. Drake, J.W. and Holland, J.J. (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A*, **96**, 13910-13913.
44. Lauring, A.S., Frydman, J. and Andino, R. (2013) The role of mutational robustness in RNA virus evolution. *Nat Rev Microbiol*, **11**, 327-336.
45. Holmes, E.C. (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol*, **11**, 543-546.
46. Perales, C. and Domingo, E. (2016) Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. *Curr Top Microbiol Immunol*, **392**, 323-339.
47. Steinhauer, D.A. and Holland, J.J. (1987) Rapid evolution of RNA viruses. *Annu Rev Microbiol*, **41**, 409-433.
48. Fitzsimmons, W.J., Woods, R.J., McCrone, J.T., Woodman, A., Arnold, J.J., Yennawar, M., Evans, R., Cameron, C.E. and Lauring, A.S. (2018) A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol*, **16**, e2006459.
49. Domingo, E. and Perales, C. (2019) Viral quasispecies. *PLoS Genet*, **15**, e1008271.
50. Domingo, E. and Schuster, P. (2016) What Is a Quasispecies? Historical Origins and Current Scope. *Curr Top Microbiol Immunol*, **392**, 1-22.
51. Crotty, S., Cameron, C.E. and Andino, R. (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A*, **98**, 6895-6900.
52. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.T., Meng, X., Xia, H., Wood, J.M., Harris, L.D. *et al.* (2021) Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife*, **10**.
53. Matsumura, E.E., Nerva, L., Nigg, J.C., Falk, B.W. and Nouri, S. (2016) Complete Genome Sequence of the Largest Known Flavi-Like Virus, Diaphorina citri flavi-

- like virus, a Novel Virus of the Asian Citrus Psyllid, *Diaphorina citri*. *Genome Announc*, **4**.
54. Debat, H. and Bejerman, N. (2023) Two novel flavivirus-like viruses shed light on the plant-infecting kosloviruses. *Arch Virol*, **168**, 184.
 55. Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.
 56. Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. and Howard, M.T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, **332**, 498-510.
 57. Plant, E.P. and Dinman, J.D. (2008) The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci*, **13**, 4873-4881.
 58. Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res*, **96**, 59-126.
 59. Angelini, M.M., Akhlaghpour, M., Neuman, B.W. and Buchmeier, M.J. (2013) Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, **4**.
 60. Oudshoorn, D., Rijs, K., Limpens, R., Groen, K., Koster, A.J., Snijder, E.J., Kikkert, M. and Bárcena, 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**.
 61. Knoops, K., Kikkert, M., Worm, S.H., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M. and Snijder, E.J. (2008) SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol*, **6**, e226.
 62. van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog*, **4**, e1000054.
 63. Wolff, G., Limpens, R., Zevenhoven-Dobbe, J.C., Laugks, U., Zheng, S., de Jong, A.W.M., Koning, R.I., Agard, D.A., Grünwald, K., Koster, A.J. *et al.* (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*, **369**, 1395-1398.
 64. Gosert, R., Kanjanahaluethai, A., Egger, D., Bienz, K. and Baker, S.C. (2002) RNA replication of mouse hepatitis virus takes place at double-membrane vesicles. *J Virol*, **76**, 3697-3708.

65. Wolff, G., Melia, C.E., Snijder, E.J. and Barcena, M. (2020) Double-Membrane Vesicles as Platforms for Viral Replication. *Trends Microbiol*, **28**, 1022-1033.
66. Malone, B., Urakova, N., Snijder, E.J. and Campbell, E.A. (2021) Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*.
67. Sola, I., Almazan, F., Zuniga, S. and Enjuanes, L. (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol*, **2**, 265-288.
68. Dufour, D., Mateos-Gomez, P.A., Enjuanes, L., Gallego, J. and Sola, I. (2011) Structure and functional relevance of a transcription-regulating sequence involved in coronavirus discontinuous RNA synthesis. *J Virol*, **85**, 4963-4973.
69. te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H. and Snijder, E.J. (2010) The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res*, **38**, 203-214.
70. 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. *et al.* (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.
71. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2020) Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell*, **184**, 184-193.
72. Wang, B., Svetlov, D. and Artsimovitch, I. (2021) NMPylation and de-NMPylation of SARS-CoV-2 nsp9 by the NiRAN domain. *Nucleic Acids Res*, **49**, 8822-8835.
73. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
74. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.
75. Anderson, T.K., Hoferle, P.J., Chojnacki, K.J., Lee, K.W., Coon, J.J. and Kirchdoerfer, R.N. (2024) An alphacoronavirus polymerase structure reveals conserved replication factor functions. *Nucleic Acids Res*.

76. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
77. Krichel, B., Falke, S., Hilgenfeld, R., Redecke, L. and Uetrecht, C. (2020) Processing of the SARS-CoV pp1a/ab nsp7-10 region. *Biochem J*, **477**, 1009-1019.
78. Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G. and 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.
79. Zhai, Y., Sun, F., Li, X., Pang, H., Xu, X., Bartlam, M. and 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.
80. Biswal, M., Diggs, S., Xu, D., Khudaverdyan, N., Lu, J., Fang, J., Blaha, G., Hai, R. and Song, J. (2021) Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res*, **49**, 5956-5966.
81. Krichel, B., Bylapudi, G., Schmidt, C., Blanchet, C., Schubert, R., Brings, L., Koehler, M., Zenobi, R., Svergun, D., Lorenzen, K. *et al.* (2021) Hallmarks of Alpha- and Betacoronavirus non-structural protein 7+8 complexes. *Sci Adv*, **7**.
82. Deming, D.J., Graham, R.L., Denison, M.R. and Baric, R.S. (2006) MHV-A59 ORF1a replicase protein nsp7-nsp10 processing in replication. *Adv Exp Med Biol*, **581**, 101-104.
83. Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C. *et al.* (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*, **182**, 417-428 e413.
84. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
85. Yuan, S., Balaji, S., Lomakin, I.B. and Xiong, Y. (2021) Coronavirus Nsp1: Immune Response Suppression and Protein Expression Inhibition. *Front Microbiol*, **12**, 752214.
86. Thoms, M., Buschauer, R., Ameisemeier, M., Koepke, L., Denk, T., Hirschenberger, M., Kratzat, H., Hayn, M., Mackens-Kiani, T., Cheng, J. *et al.* (2020) Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science*, **369**, 1249-1255.
87. Jiao, Y., Zhao, P., Xu, L.D., Yu, J.Q., Cai, H.L., Zhang, C., Tong, C., Yang, Y.L., Xu, P., Sun, Q. *et al.* (2024) Enteric coronavirus nsp2 is a virulence determinant that

- recruits NBR1 for autophagic targeting of TBK1 to diminish the innate immune response. *Autophagy*, 1-18.
88. Graham, R.L., Sims, A.C., Baric, R.S. and Denison, M.R. (2006) The nsp2 proteins of mouse hepatitis virus and SARS coronavirus are dispensable for viral replication. *Adv Exp Med Biol*, **581**, 67-72.
 89. Xu, Z., Choi, J.H., Dai, D.L., Luo, J., Ladak, R.J., Li, Q., Wang, Y., Zhang, C., Wiebe, S., Liu, A.C.H. *et al.* (2022) SARS-CoV-2 impairs interferon production via NSP2-induced repression of mRNA translation. *Proc Natl Acad Sci U S A*, **119**, e2204539119.
 90. Lei, J., Kusov, Y. and Hilgenfeld, R. (2018) Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral Res*, **149**, 58-74.
 91. Zimmermann, L., Zhao, X., Makroczyova, J., Wachsmuth-Melm, M., Prasad, V., Hensel, Z., Bartenschlager, R. and Chlanda, P. (2023) SARS-CoV-2 nsp3 and nsp4 are minimal constituents of a pore spanning replication organelle. *Nat Commun*, **14**, 7894.
 92. Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., Becker, S., Rox, K. and Hilgenfeld, R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*, **368**, 409-412.
 93. Lu, Y., Lu, X. and Denison, M.R. (1995) Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. *J Virol*, **69**, 3554-3559.
 94. Oostra, M., Hagemeijer, M.C., van Gent, M., Bekker, C.P., te Lintelo, E.G., Rottier, P.J. and 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.
 95. Park, G.J., Osinski, A., Hernandez, G., Eitson, J.L., Majumdar, A., Tonelli, M., Henzler-Wildman, K., Pawłowski, K., Chen, Z., Li, Y. *et al.* (2022) The mechanism of RNA capping by SARS-CoV-2. *Nature*, **609**, 793-800.
 96. Bouvet, M., Lugari, A., Posthuma, C.C., Zevenhoven, J.C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J.C., Lecine, P. *et al.* (2014) Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J Biol Chem*, **289**, 25783-25796.
 97. Chan, J.F., Kok, K.H., Zhu, Z., Chu, H., To, K.K., Yuan, S. and Yuen, K.Y. (2020) Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg Microbes Infect*, **9**, 221-236.

98. Ivanov, K.A. and 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.
99. Ivanov, K.A., Thiel, V., Dobbe, J.C., van der Meer, Y., Snijder, E.J. and Ziebuhr, J. (2004) Multiple enzymatic activities associated with severe acute respiratory syndrome coronavirus helicase. *J Virol*, **78**, 5619-5632.
100. Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B. *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog*, **6**, e1000896.
101. Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B. and 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.
102. Bhardwaj, K., Sun, J., Holzenburg, A., Guarino, L.A. and Kao, C.C. (2006) RNA recognition and cleavage by the SARS coronavirus endoribonuclease. *J Mol Biol*, **361**, 243-256.
103. Decroly, E., Imbert, I., Coutard, B., Bouvet, M., Selisko, B., Alvarez, K., Gorbalenya, A.E., Snijder, E.J. and Canard, B. (2008) Coronavirus nonstructural protein 16 is a cap-0 binding enzyme possessing (nucleoside-2'O)-methyltransferase activity. *J Virol*, **82**, 8071-8084.
104. Lehmann, K.C., Snijder, E.J., Posthuma, C.C. and Gorbalenya, A.E. (2015) What we know but do not understand about nidovirus helicases. *Virus Res*, **202**, 12-32.
105. Zhang, R., Li, Y., Cowley, T.J., Steinbrenner, A.D., Phillips, J.M., Yount, B.L., Baric, R.S. and Weiss, S.R. (2015) The nsp1, nsp13, and M proteins contribute to the hepatotropism of murine coronavirus JHM.WU. *J Virol*, **89**, 3598-3609.
106. Deng, X., Hackbart, M., Mettelman, R.C., O'Brien, A., Mielech, A.M., Yi, G., Kao, C.C. and Baker, S.C. (2017) Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc Natl Acad Sci U S A*, **114**, E4251-e4260.
107. Hackbart, M., Deng, X. and Baker, S.C. (2020) Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc Natl Acad Sci U S A*, **117**, 8094-8103.
108. Ogando, N.S., El Kazzi, P., Zevenhoven-Dobbe, J.C., Bontes, B.W., Decombe, A., Posthuma, C.C., Thiel, V., Canard, B., Ferron, F., Decroly, E. and Snijder, E.J. (2021) Structure–function analysis of the nsp14 N7–guanine methyltransferase

- reveals an essential role in Betacoronavirus replication. *Proceedings of the National Academy of Sciences*, **118**, e2108709118.
109. Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D. and Baric, R.S. (2011) Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol*, **8**, 270-279.
 110. Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M. and Denison, M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog*, **9**, e1003565.
 111. Ferron, F., Subissi, L., Silveira De Moraes, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B. and 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.
 112. Matsuda, A., Plewka, J., Rawski, M., Mourão, A., Zajko, W., Siebenmorgen, T., Kresik, L., Lis, K., Jones, A.N., Pachota, M. *et al.* (2024) Despite the odds: formation of the SARS-CoV-2 methylation complex. *Nucleic Acids Res*.
 113. Černý, J., Černá Bolfíková, B., Valdés, J.J., Grubhoffer, L. and Růžek, D. (2014) Evolution of tertiary structure of viral RNA dependent polymerases. *PLoS One*, **9**, e96070.
 114. Peersen, O.B. (2017) Picornaviral polymerase structure, function, and fidelity modulation. *Virus Res*, **234**, 4-20.
 115. Malone, B.F., Perry, J.K., Olinares, P.D.B., Lee, H.W., Chen, J., Appleby, T.C., Feng, J.Y., Bilello, J.P., Ng, H., Sotiris, J. *et al.* (2023) Structural basis for substrate selection by the SARS-CoV-2 replicase. *Nature*, 1-7.
 116. Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *PNAS*, **107**.
 117. Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S. and Vassilyev, D.G. (2004) Structural basis for substrate selection by t7 RNA polymerase. *Cell*, **116**, 381-391.
 118. Shu, B. and Gong, P. (2016) Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *PNAS*, **113**.
 119. Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M. *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun*, **11**, 5874.

120. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T. *et al.* (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
121. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, **118**.
122. Cheung, A.C. and Cramer, P. (2011) Structural basis of RNA polymerase II backtracking, arrest and reactivation. *Nature*, **471**, 249-253.
123. Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M. and Kornberg, R.D. (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*, **324**, 1203-1206.
124. Abdelkareem, M., Saint-André, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J. and Weixlbaumer, A. (2019) Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell*, **75**, 298-309.e294.
125. Bębenek, A. and Ziuzia-Graczyk, I. (2018) Fidelity of DNA replication-a matter of proofreading. *Curr Genet*, **64**, 985-996.
126. Buchel, G., Nayak, A.R., Herbine, K., Sarfallah, A., Sokolova, V.O., Zamudio-Ochoa, A. and Temiakov, D. (2023) Structural basis for DNA proofreading. *Nat Commun*, **14**.
127. Yan, L., Yang, Y., Li, M., Zhang, Y., Zheng, L., Ge, J., Huang, Y.C., Liu, Z., Wang, T., Gao, S. *et al.* (2021) Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell*, **184**, 3474-3485.e3411.
128. Perry, J.K., Appleby, T.C., Bilello, J.P., Feng, J.Y., Schmitz, U. and Campbell, E.A. (2021) An atomistic model of the coronavirus replication-transcription complex as a hexamer assembled around nsp15. *J Biol Chem*, **297**, 101218.
129. Zuo, Y. and Deutscher, M.P. (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res*, **29**, 1017-1026.
130. Baddock, H.T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L.P., Cruz-Migoni, A., Fan, H., Keown, J.R., Walker, A.P. *et al.* (2022) Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Res*, **50**, 1484-1500.

131. Ma, Z., Pourfarjam, Y. and Kim, I.K. (2021) Reconstitution and functional characterization of SARS-CoV-2 proofreading complex.
132. Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. and 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.
133. Ogando, N.S., Zevenhoven-Dobbe, J.C., van der Meer, Y., Bredenbeek, P.J., Posthuma, C.C. and Snijder, E.J. (2020) The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol*.
134. Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L. and Denison, M.R. (2007) High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol*, **81**, 12135-12144.
135. Smith, E.C., Case, J.B., Blanc, H., Isakov, O., Shomron, N., Vignuzzi, M. and Denison, M.R. (2015) Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity. *J Virol*, **89**, 6418-6426.
136. Fidalgo da Silva, E. and Reha-Krantz, L.J. (2007) DNA polymerase proofreading: active site switching catalyzed by the bacteriophage T4 DNA polymerase. *Nucleic Acids Res*, **35**, 5452-5463.
137. Pruijssers, A.J., George, A.S., Schäfer, A., Leist, S.R., Gralinski, L.E., Dinnon, K.H., 3rd, Yount, B.L., Agostini, M.L., Stevens, L.J., Chappell, J.D. *et al.* (2020) Remdesivir Inhibits SARS-CoV-2 in Human Lung Cells and Chimeric SARS-CoV Expressing the SARS-CoV-2 RNA Polymerase in Mice. *Cell Reports*, **32**.
138. Yan, V.A.-O. and Muller, F.L. (2021) Why Remdesivir Failed: Preclinical Assumptions Overestimate the Clinical Efficacy of Remdesivir for COVID-19 and Ebola. *Antimicrobial Agents and Chemotherapy*, **65**.
139. Beigel, J., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S. *et al.* (2020) Remdesivir for the Treatment of Covid-19 - Final Report. *The New England Journal of Medicine*, **383**.
140. Agostini, M.L., Pruijssers, A.J., Chappell, J.D., Gribble, J., Lu, X., Andres, E.L., Bluemling, G.R., Lockwood, M.A., Sheahan, T.P., Sims, A.C. *et al.* (2019) Small-Molecule Antiviral β -d-N (4)-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J Virol*, **93**.
141. Wahl, A., Gralinski, L.E., Johnson, C.E., Yao, W., Kovarova, M., Dinnon, K.H., 3rd, Liu, H., Madden, V.J., Krzystek, H.M., De, C. *et al.* (2021) SARS-CoV-2 infection is effectively treated and prevented by EIDD-2801. *Nature*, **591**, 451-457.

142. Cox, R.M., Wolf, J.D. and Plemper, R.K. (2021) Therapeutically administered ribonucleoside analogue MK-4482/EIDD-2801 blocks SARS-CoV-2 transmission in ferrets. *Nat Microbiol*, **6**, 11-18.
143. Jayk Bernal, A., Gomes da Silva, M.M., Musungaie, D.B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., Martín-Quirós, A., Caraco, Y., Williams-Diaz, A., Brown, M.L. *et al.* (2022) Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *N Engl J Med*, **386**, 509-520.
144. Saravolatz, L.D., Depcinski, S. and Sharma, M. (2023) Molnupiravir and Nirmatrelvir-Ritonavir: Oral Coronavirus Disease 2019 Antiviral Drugs. *Clin Infect Dis*, **76**, 165-171.
145. Butler, C.C., Hobbs, F.D.R., Gbinigie, O.A., Rahman, N.M., Hayward, G., Richards, D.B., Dorward, J., Lowe, D.M., Standing, J.F., Breuer, J. *et al.* (2023) Molnupiravir plus usual care versus usual care alone as early treatment for adults with COVID-19 at increased risk of adverse outcomes (PANORAMIC): an open-label, platform-adaptive randomised controlled trial. *Lancet*, **401**, 281-293.
146. Owen, D.R., Allerton, C.M.N., Anderson, A.S., Aschenbrenner, L., Avery, M., Berritt, S., Boras, B., Cardin, R.D., Carlo, A., Coffman, K.J. *et al.* (2021) An oral SARS-CoV-2 M(pro) inhibitor clinical candidate for the treatment of COVID-19. *Science*, **374**, 1586-1593.
147. Hammond, J., Leister-Tebbe, H., Gardner, A., Abreu, P., Bao, W., Wisemandle, W., Baniecki, M., Hendrick, V.M., Damle, B., Simón-Campos, A. *et al.* (2022) Oral Nirmatrelvir for High-Risk, Nonhospitalized Adults with Covid-19. *N Engl J Med*, **386**, 1397-1408.
148. Kokic, G., Hillen, H.S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Hobartner, C. and Cramer, P. (2021) Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun*, **12**, 279.
149. Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kokic, G., Hillen, H.S., Höbartner, C. and Cramer, P. (2021) Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol*, **28**, 740-746.

Chapter 2: An alphacoronavirus polymerase structure reveals conserved replication factor functions

Authors: Thomas K. Anderson^{1,2,3}, Peter J. Hoferle^{1,2,3}, Kennan J. Chojnacki^{1,2,3}, Kenneth W. Lee⁴, Joshua J. Coon^{4,5,6}, Robert N. Kirchdoerfer^{1,2,3*}

Affiliations:

¹Biochemistry Department, University of Wisconsin-Madison, Madison, WI 53706 United States. ²Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI 53706, United States. ³Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, United States. ⁴Biomolecular Chemistry Department, University of Wisconsin-Madison, Madison, WI 53706, United States. ⁵Department of Chemistry, University of Wisconsin-Madison, Madison, Wisconsin 53715, United States. ⁶Morgridge Institute for Research, Madison, Wisconsin 53715, United States.

Published as:

Thomas K Anderson, Peter J Hoferle, Kennan J Chojnacki, Kenneth W Lee, Joshua J Coon, Robert N Kirchdoerfer, An alphacoronavirus polymerase structure reveals conserved replication factor functions, *Nucleic Acids Research*, Volume 52, Issue 10, 10 June 2024, Pages 5975–5986, <https://doi.org/10.1093/nar/gkae153>

Author contributions:

T.K.A. produced the data for all Figures in this manuscript.

K.W.L. produced the data for Table 1.

T.K.A. and R.N.K. developed this project

T.K.A. wrote the manuscript with R.N.K.

K.W.L. and J.J.C. provided comments and edits on the manuscript

Abstract:

Coronaviruses are a diverse subfamily of viruses containing pathogens of humans and animals. This subfamily of viruses replicates their RNA genomes using a core polymerase complex composed of viral non-structural proteins: nsp7, nsp8 and nsp12. Most of our understanding of coronavirus molecular biology comes from betacoronaviruses like SARS-CoV and SARS-CoV-2, the latter of which is the causative agent of COVID-19. In contrast, members of the alphacoronavirus genus are relatively understudied despite their importance in human and animal health. Here we have used cryo-electron microscopy to determine structures of the alphacoronavirus porcine epidemic diarrhea virus (PEDV) core polymerase complex bound to RNA. One structure shows an unexpected nsp8 stoichiometry despite remaining bound to RNA. Biochemical analysis shows that the N-terminal extension of one nsp8 is not required for *in vitro* RNA synthesis for alpha and betacoronaviruses. Our work demonstrates the importance of studying diverse coronaviruses in revealing aspects of coronavirus replication and identifying areas of conservation to be targeted by antiviral drugs.

Introduction:

The coronavirus (CoV) subfamily is composed of four genera: the *Alpha*-, *Beta*-, *Gamma*-, and *Delta*CoVs. Across all these genera there are diverse human and animal pathogens that cause a wide range of disease severities (1). Since 2002, three betaCoVs have emerged from animal reservoirs and caused human epidemics or pandemics: SARS-CoV, MERS-CoV, and SARS-CoV-2 (2-5). The most recent, SARS-CoV-2, emerged in 2019 and is the causative agent of the COVID-19 pandemic (5,6). There are four endemic common-cold causing human CoVs: of these there are two alphaCoVs, HuCoV-229E and HuCoV-NL63, and two betaCoVs, HuCoV-HKU1 and HuCoV-OC43 (1). In 2021, a recombined feline/canine *AlphaCoV*, named CCoV-HuPn-2018, was identified in samples from human patients with pneumonia in Malaysia and the United States (7-9). CCoV-HuPn-2018 is the only recorded human emergent *AlphaCoV* in recent years and exemplifies the persistent threat of *AlphaCoV* spillover. In the *AlphaCoV* genera there are several porcine pathogens including transmissible gastroenteritis virus (TGEV), swine enteric alphacoronavirus (SeACoV), and porcine epidemic diarrhea virus (PEDV) (10,11). Since 2010, PEDV has had detrimental impacts on the global swine industry, with outbreaks having 70-100% unweaned piglet fatality rates at swine farms despite vaccination efforts (12-14).

CoVs are positive-sense, single-stranded RNA viruses with large, ~30 kilobase genomes (1,15). To replicate their genomes CoVs encode an RNA-dependent RNA polymerase (RdRP), termed non-structural protein 12 (nsp12) (16-18). Nsp12 also contains the nidovirus RdRP-associated nucleotidyltransferase (NiRAN) domain at its N terminus, which has been shown to be involved in mRNA capping (19,20). *In vitro* studies

of betaCoVs have shown that viral replication factors nsp7 and nsp8 interact with nsp12 to form an active and processive polymerase complex (21). The SARS-CoV-2 complex is the fastest known RdRP with nucleotide addition rates up to 170 nt/sec *in vitro* (22). The first structure of a CoV polymerase complex from the *BetaCoV*, SARS-CoV, revealed a subunit stoichiometry of one nsp12, two nsp8, and one nsp7 where nsp7 and one nsp8 form a heterodimer (19). The nsp8 that directly interacts with nsp12 is denoted as nsp8_F (fingers), and the nsp8 that interacts with nsp7 and nsp12 is denoted as nsp8_T (thumb). In this paper we refer to this complex of nsp12, nsp7, and nsp8 as the CoV core polymerase complex.

Recent SARS-CoV-2 core polymerase structures have provided key insights into *BetaCoV* replication. Nearly all structures of SARS-CoV-2 complexes shown to be enzymatically active or bind RNA *in vitro* had the same 1:2:1 nsp7:nsp8:nsp12 stoichiometry as SARS-CoV (23,24). An exception is a cryo-EM structure of the SARS-CoV-2 polymerase complex stalled by Remdesivir that remained bound to RNA but lacked any density for nsp8_T (25). Several SARS-CoV-2 polymerase structures determined with longer dsRNA substrates have shown each nsp8's N-terminal extension bound to upstream dsRNA as it exits the polymerase active site, which has been hypothesized to promote processivity through a "sliding-pole" mechanism (23,24). Addition of the viral RNA helicase, nsp13, to core polymerase complexes showed that each nsp8 subunit independently scaffolds a nsp13 (two nsp13s bind a single core complex) (20,26,27). The nsp13 associated with nsp8_T was shown to be capable of binding to the 5' end of template RNA prior to the RNA entering the polymerase active site (27). Nsp13's association with the template RNA produced a conundrum, where

nsp12 and nsp13 are oriented such that the polymerase and helicase translocate in opposite directions. Further biochemical and structural work demonstrated that this nsp13 can stimulate template backtracking of the polymerase, during which the 3' end of nascent RNA extrudes through the nsp12 NTP channel (28). While this nsp13 has been implicated in backtracking, the triggers and regulation of backtracking are unknown.

To date, there is limited knowledge of CoV polymerase biochemistry and structural biology outside the *Sarbecovirus* subgenus (SARS-CoV and SARS-CoV-2). To our knowledge, *in vitro* demonstration of robust nsp12 polymerase activity has yet to be demonstrated for an *Alpha*-, *Gamma*-, or *Delta*CoV. Similarly, outside of betaCoVs there is limited structural and biochemical information for nsp7 and nsp8. A crystal structure of the *Alpha*CoV feline coronavirus (FCoV) nsp7-nsp8 complex revealed that *Alpha*- and *Beta*CoV nsp7s and the nsp8 head domains have high structural homology (29). Previous native mass spectrometry experiments of alpha- and betaCoV nsp7-nsp8 complexes determined that each genera's cofactors form homo- and hetero-oligomers, although the oligomers sampled by each genus varied in cofactor ratio and/or prevalence of particular oligomers (30). Across CoV genera there is sequence conservation (>40%) of the nsps involved in RNA synthesis (i.e., nsp7, nsp8, and nsp12) (31). Sequence and available structural homology suggest shared mechanisms among these viruses to replicate their RNA genomes. A recent publication showed that SARS-CoV-2 replication complexes assembled with either a truncated nsp8_F or nsp8_T, lacking their respective RNA binding domains, were capable of processive RNA synthesis similar to the wildtype complex (32). These results challenge the aforementioned "sliding-pole" hypothesis, further displaying

the need for more biochemical and structural characterization of CoV polymerases and their replication factors.

Though our knowledge of SARS-CoV and SARS-CoV-2 replication mechanics has improved over the last two decades, the narrow structural biology focus on two closely related betaCoVs limits our understanding of CoV replication across the virus subfamily. To address this gap, we studied the polymerase core complex of the *AlphaCoV* PEDV using biochemistry and structural biology. Our structure of the PEDV core polymerase complex bound to an RNA primer-template pair without nsp8_T shows the lability of nsp8_T to participate in this complex and additional mutagenesis and biochemistry demonstrate that the nsp8_T N-terminal helical extension is not necessary for either *Alpha-* or *BetaCoV* RNA synthesis activity *in vitro*. The identification of conserved mechanisms and structural motifs between alpha- and betaCoVs will allow for the development of broadly acting CoV therapeutic strategies.

Materials and Methods:

Expression Construct Design

DNA encoding PEDV and SARS-CoV-2 nsps were codon optimized and synthesized (Genscript). PEDV protein sequences correspond to GenBank AKJ21892.1. SARS-CoV-2 protein sequences correspond to GenBank UHD90671.1. PEDV nsp7 and nsp8 genes were cloned into pET46 and pET45b expression vectors, respectively. PEDV nsp7 was cloned with a C-terminal TEV protease site and hexahistidine tag. PEDV nsp8 has an N-terminal hexahistidine tag and TEV protease site. SARS-CoV-2 nsp7 and nsp8 were both cloned into pET46 with N-terminal hexahistidine tags, and enterokinase and

TEV protease sites. Both PEDV and SARS-CoV-2 nsp12 genes were cloned into pFastBac with C-terminal TEV protease sites and double Strep II tags.

Expression plasmids for nsp12 point mutants were produced using mutagenesis on the pFastBac plasmids. The nsp8-nsp7 (nsp8L7) fusion constructs were produced by overlap PCR creating a GSGSGS peptide linker between nsp8 and nsp7 and inserted into pET46 expression vectors with N-terminal hexahistidine tags, and enterokinase and TEV protease sites. Truncation constructs were produced using Kinase-Ligase-DpnI (KLD) cloning. All open reading frames on the DNA plasmids were verified by Sanger sequencing.

Recombinant Protein Expression

Recombinant nsp7 and nsp8: Replication factor proteins were expressed in Rosetta 2pLysS *E. coli* cells (Novagen). Cultures were grown at 37°C and induced at an OD₆₀₀ of 0.6-0.8 with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 500 μM. After growing for 16 hours at 16°C, cells were harvested by centrifugation and resuspended in replication factor wash buffer (10 mM Tris-Cl pH 8.0, 300 mM sodium chloride, 30 mM imidazole, and 2 mM dithiothreitol (DTT)). Cells were lysed in a microfluidizer (Microfluidics) and lysates cleared via centrifugation. Replication factors were purified using Ni-NTA agarose beads (Qiagen), eluting with 300 mM imidazole. Eluted protein was digested with 1% (w/w) TEV protease overnight while dialyzing (10 mM Tris-Cl pH 8.0, 300 mM sodium chloride, and 2 mM DTT) at 4°C. Digested protein was flowed back over Ni-NTA agarose beads to removed undigested protein, and further purified using a Superdex 200 Increase 10/300 GL column (Cytiva) in 25 mM Tris-Cl pH 8.0, 300 mM sodium chloride, and 2 mM DTT. Fractions containing the

protein of interest were concentrated using ultrafiltration. Concentrated protein was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. Replication factor protein yields for 1 L of cells was 10-40 mg.

Recombinant nsp12s: pFastBac plasmids carrying the nsp12 gene and DH10Bac *E. coli* (Life Technologies) were used to produce recombinant bacmids for each gene. Bacmids were transfected into Sf9 cells (Expression Systems) with Cellfectin II (Life Technologies) to produce recombinant baculoviruses, which were twice amplified using Sf9 cells. Amplified baculoviruses were used to infect Sf21 cells (Expression Systems) for protein expression. After two days of incubation at 27°C, cells were collected via centrifugation and pellets resuspended in wash buffer (25 mM HEPES pH 7.4, 300 mM sodium chloride, 1 mM magnesium chloride, and 5 mM DTT) with an added 143 μ L of BioLock (IBA) per 1 L of culture. Cells were lysed using a microfluidizer (Microfluidics) and lysates cleared via centrifugation. Protein was affinity purified using Streptactin superflow agarose (IBA) and eluted with wash buffer that contained 2.5 mM desthiobiotin. Protein was further purified via size exclusion chromatography with a Superdex 200 Increase 10/300 GL column (Cytiva) in 25 mM HEPES pH 7.4, 300 mM sodium chloride, 100 μ M magnesium chloride, and 2 mM tris(2-carboxyethyl)phosphine (TCEP). Fractions containing nsp12 were pooled and concentrated using ultrafiltration. Concentrated protein was aliquoted, flash frozen in liquid nitrogen and stored at -80°C. Average protein yield for 1 L of culture was 3-5 mg.

Preparation of RNA Substrates

RNA oligos were purchased from Integrated DNA Technologies. Primer RNAs were modified with a 5' fluorescein to monitor the RNA by gel electrophoresis.

Pair 1:

Primer RNA: CAUUCUCCUAAGAAGCUAUUAAAAUCACA

Template RNA: AAAAAGGGUUGUGAUUUUAAUAGCUUCUUAGGAGAAUG.

Pair 2:

Primer RNA: CAUUCUCCUAAGAAGCUAUUAAAUCACAGAUU

Template RNA:

CAGUGUCAUGGAAAAACAGAAAAAUCUGUGAUUUUAAUAGCUUCUUAGGAGAAUG

RNA template was always held in slight excess of primer (primer : template ratio of 1:1.2). RNA oligos were annealed in 2.5 mM HEPES pH 7.4, 2.5 mM potassium chloride, and 0.5 mM magnesium chloride and heated at 95°C for 5 minutes, then allowed to cool slowly back to 25°C for 75 minutes before either being used immediately or stored at -20°C.

Native Mass Spectrometry

Component concentrations were, unless stated otherwise, 10.4 μ M nsp12, 20.8 μ M nsp7, 31.2 μ M nsp8 (or nsp8-nsp7 fusion), and 12.5 μ M RNA duplex (Pair 1). Proteins were combined in native-MS buffer (10 mM Tris-Cl pH 8.0, 100 mM ammonium acetate, 2 mM magnesium chloride, and 1 mM DTT) and incubated at 25°C for 15 minutes. RNA duplex was then added, and reactions were incubated at 25°C for an additional 15 minutes. SARS-CoV-2 T853R and PEDV V848R complexes (homologous mutations) were prepared at half the normal protein and RNA concentrations due to low yields of the mutant nsp12 proteins.

All samples were buffer exchanged into 100 mM ammonium acetate (NH_4OAc) with 2 mM MgCl_2 . First, 350 μL 100 mM NH_4OAc with 2 mM MgCl_2 was added to a 100 kDa molecular weight cut-off Amicon Ultra-0.5 Centrifugal Filter Unit (Millipore Sigma), followed by 15 μL of ~ 10 μM polymerase sample. The samples were centrifuged at $14,000 \times g$ for 10 min to load the sample, followed by two rinses with 400 μL of the $\text{NH}_4\text{OAc}/\text{MgCl}_2$ solution, also centrifuged at $14,000 \times g$ for 10 min. Finally, the filter was inverted into a new catch tube and centrifuged at $1,000 \times g$ for 2 min to recover the filtered and concentrated sample. Samples were diluted using the same $\text{NH}_4\text{OAc}/\text{MgCl}_2$ solution to roughly 1–5 μM for introduction to the mass spectrometer via electrospray ionization.

Standard wall borosilicate tubing (1.20 mm o.d., 0.69 mm i.d., Sutter Instrument) was pulled using a P-1000 Micropipette Puller (Sutter Instrument) to a tapered tip of ~ 3 –5 μm diameter. Each sample was loaded into a pulled glass capillary and 1.1–1.3 kV was applied to the sample using a platinum wire inserted into the back of the capillary. Charged droplets entered a Q Exactive UHMR Hybrid Quadrupole Orbitrap Mass Spectrometer (ThermoFisher Scientific) via a heated inlet capillary. The heated capillary was set to 200–250°C to minimize solvent adduction to analyte ions. Additional removal of adducts was accomplished using in-source trapping with a range of injection voltages typically between 150–300 V. The voltage was tuned for each sample to maximize adduct removal with minimal dissociation of the ionized complex. For these experiments the vacuum was set to a value of 10 resulting in a pressure readout of $2e-4$ mbar and the S-lens radiofrequency level was set to a value of 200. All reported spectra are an average of 50 scans collected with the Orbitrap mass analyzer at a resolution setting of 6,250. Charge states were assigned manually by minimizing the standard deviation of masses

calculated by peaks within a single charge state distribution. The average mass and standard deviation are reported for major distributions in each spectrum.

In vitro Primer Extension Assay

Assay conditions were 10 mM Tris-Cl pH 8.0, 10 mM sodium chloride, 2 mM magnesium chloride, and 1 mM DTT with a typical reaction volume of 20 μ L. Protein final concentrations were 500 nM nsp12, 1.5 μ M nsp7, and 1.5 μ M nsp8 (nsp8-nsp7 fusion and truncation proteins were also at 1.5 μ M). Duplex RNA (Pair 1) final concentration was 250 nM. Prior to use, proteins were diluted in assay buffer. Diluted proteins were then combined and incubated at 25°C for 15 minutes, duplex RNA was added and reactions incubated at 25°C for an additional 15 minutes. Reactions were initiated by the addition of NTPs to a final concentration of 40 μ M and reactions ran for 1 minute (at 25°C for SARS-CoV-2 polymerases, or 30°C for PEDV polymerases) before being halted by addition of two volumes of sample loading buffer (95% (v/v) formamide, 2 mM ethylenediaminetetraacetic acid (EDTA), and 0.75 mM bromophenol blue). Samples were heated at 95°C then analyzed using denaturing urea-PAGE (8 M urea, 15% polyacrylamide) run in 1X TBE (89 mM Tris-Cl pH 8.3, 89 mM boric acid, 2 mM EDTA). Gels were imaged using a Typhoon FLA 9000 (GE Healthcare) to identify fluorescein signals. Extension was quantified using ImageJ (33).

Sample Preparation and Grid Freezing for CryoEM

For the complex lacking nsp8_T:

PEDV polymerase complexes were prepared at a total protein concentration of 1 mg/mL with a ratio of 2 : 2 : 1 : 1.2 nsp7 : nsp8 : nsp12 : RNA duplex (Pair 1). The complex

was assembled in 25 mM HEPES pH 7.5, 50 mM sodium chloride, 2 mM magnesium chloride, and 2 mM DTT. Proteins were diluted in buffer then immediately combined and incubated at 25°C for 15 minutes before RNA was added and incubated at 25°C for another 15 minutes. Samples were stored on ice prior to grid freezing.

Samples were prepared for structural analysis using UltraAuFoil R1.2/1.3 300 mesh grids (Quantifoil) and a Vitrobot Mark IV (ThermoFisher Scientific). Grids were freshly glow discharged using a GloQube Plus (Quorum) for 20 seconds with a current of 20 mA in an air atmosphere, creating a negative charge on the grid surface. Immediately before applying sample to grids, 0.5 μ L n-dodecyl- β -D-maltoside (DDM) was added to samples at a final concentration of 60 μ M. 3.5 μ L of sample was spotted onto grids before double-sided blotting and plunge freezing in liquid ethane. Samples were blotted for 4 seconds at a blot force of -15 with chamber conditions of 100% humidity and 4°C.

For the complete complex with nsp8_T:

PEDV polymerase complex were prepared at a total protein concentration of 1 mg/mL with a ratio of 2.2 : 2.2 : 1 : 1.2 nsp7 : nsp8 : nsp12 : RNA duplex (Pair 2). The complex was assembled using the same buffer and protocol as above.

Samples were frozen on Quantifoil R1.2/1.3 200 mesh grids using a Vitrobot Mark IV (ThermoFisher Scientific). Grids were freshly glow discharged as described above. Right before blotting, 0.5 μ L of 3-([3-Cholamidopropyl] dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO) was added to samples at a final concentration of 6 mM. 3 μ L of sample was spotted onto grids before double-sided blotting for 8 seconds at a blot force of -6 with chamber conditions of 100% humidity and 4°C. Following blotting, samples were plunge frozen in liquid ethane.

Cryo-EM Data Collection, Processing, and Model Building

For the complex lacking nsp8_T:

SerialEM was used for data collection on a Titan Krios 300 kV transmission electron microscope (ThermoFisher Scientific) (34). Movies were collected on a K3 direct electron detector (Gatan) in CDS mode with a GIF quantum energy filter slit width of 20 eV and a stage tilt of 25°. Data was collected at a magnification of 105,000x, a pixel size of 0.834 Å, and a defocus range of -0.75 to -1.75 μm at a step size of 0.5 μm.

Data processing was performed using cryoSPARC v3.3.1 (35). After patch motion correction and patch CTF estimation, particle picking was performed using blob picker with a diameter of 125-175 Å and a minimum separation between particles of 12.5 Å. Particles were extracted with a box size of 256 pixels and subjected to 2D classification with a window radius of 0.33 – 0.66 for 100 classes with 40 online-EM iterations and a batch size of 1000. Force max over poses/shifts was turned off to improve visualization of the complex during 2D classification. Three initial models were created ab-initio and further used for heterogenous refinement. Particles from two maps that looked like polymerase complexes were pooled and used for non-uniform refinement under default parameters. The 3D reconstruction was used for 3D variability analysis (36) and further 3D classification before a final non-uniform refinement and 3D reconstruction (**Table S1, Fig. S1 and S2**). Any 3D jobs run in cryoSPARC were done under default parameters unless noted otherwise.

To build the PEDV polymerase complex coordinate model, a SARS-CoV-2 polymerase complex structure (PDB ID: 7CYQ) was docked into the cryo-EM map in Coot and regions of the protein (including whole chains of nsp8, nsp9, and nsp13) and RNA

that lacked map density were removed (20,37). Mutation of protein chains to PEDV sequences, further model building, and validation was performed in Coot. The PEDV model was refined using real space refinement in Phenix (38). Refinement in Phenix and model building/validation in Coot was an iterative process to produce a coordinate model. Final model adjustments were made with ISOLDE (39).

For the complete complex with nsp8_T:

EPU was used for data collection on a Talos Arctica 200 keV transmission electron microscope (ThermoFisher Scientific). Movies were collected on a K3 direct electron detector (Gatan) in CDS mode with a GIF quantum energy filter slit width of 20 eV and no stage tilt. Data was collected at a magnification of 79,000x, a pixel size of 1.064 Å, and a defocus range -0.5 to -2.0 μm at a step size of 0.5 μm (**Table S1**).

Data was processed using cryoSPARC v4.1.0 (35). Motion correction, CTF estimation, particle picking, particles extraction, 2D classification, ab-initio model generation, 3D classification, and non-uniform refinement were done using matching parameters as above to produce the final 3D reconstruction (**Table S1, Fig. S1 and S2**). No 3D variability analysis was used in producing this 3D reconstruction.

To build the coordinate model, our initial PEDV model was docked into the cryo-EM map in ChimeraX, then we superimposed a SARS-CoV-2 polymerase complex structure (7KRP) with our PEDV model using the Matchmaker function in ChimeraX (28,40). Regions except for nsp8_T and the nsp8_F N-terminal extension were deleted from the SARS-CoV-2 model. The reduced model was saved in relation to our PEDV model in space. In Coot our PEDV model and the SARS-CoV-2 nsp8 coordinates were merged, nsp8 protein chains were mutated to PEDV sequences, and initial refinement was

performed (37). Refinement in Phenix and model building/validation in Coot was an iterative process to produce a coordinate model, the final adjustments for which were made with ISOLDE (38,39).

Figure Making

Images of protein models and maps used for figures were made in ChimeraX (40). Bar graphs were produced in Microsoft Excel. Figures were assembled in Adobe Illustrator.

Results:

Assembly of an active PEDV polymerase complex

To study the PEDV polymerase complex we recombinantly expressed and purified nsp7, nsp8, and nsp12 (**Fig. S3**). Assembly of these proteins with a short RNA duplex revealed a complex weight of 185,120 Da (+/- 70 Da) by native mass spectrometry (**Table 1, Fig. S4**). This mass is equivalent to one nsp12, one nsp7, two nsp8s, and one RNA duplex. Using an *in vitro* primer extension assay we saw polymerase activity in the presence of all three nsps (**Fig. 1A and Fig. S5**). The necessity of nsp7, nsp8 and nsp12 for robust primer extension, and the observed nsp binding stoichiometry are shared between PEDV, SARS-CoV and SARS-CoV-2 (17,19,21,23).

Sample (Expected Mass in Da)	Major Peak(s) (Da)	Major Complex(es):	Minor Peak(s) (Da)	Minor Complex(es):
PEDV nsp12 +8+7+RNA (184,293)	185,120 ± 70	Intact polymerase complex	109,450 ± 40	Free nsp12
PEDV nsp12 +8+RNA (152,529)	109,600 ± 30	Free nsp12	None	None
PEDV nsp12-A382R +8+7+RNA (184,293)	109,533 ± 60 185,120 ± 30	Free nsp12, Intact polymerase complex	218,880 ± 50	Dimeric nsp12
PEDV nsp12-V848R +8+7+RNA (184,293)	185,200 ± 200	Intact polymerase complex	370,700 ± 600 109,630 ± 200	Dimeric polymerase complex, free nsp12
PEDV nsp12 +8+8L7+RNA (183,913)	184, 670 ± 60	Intact polymerase complex	109,420 ± 30	Free nsp12
SARS-CoV-2 nsp12 +8+7+RNA (185,245)	186,240 ± 70	Intact polymerase complex	110,680 ± 20	Free nsp12
SARS-CoV-2 nsp12 +8+RNA (154,011)	110,870 ± 90	Free nsp12	221,320 ± 40	Dimeric nsp12
SARS-CoV-2 nsp12- L387R +8+7+RNA (185,245)	186,400 ± 100	Intact polymerase complex	110,900 ± 100 22,400 ± 100	Free nsp12, excess nsp8 and/or RNA
SARS-CoV-2 nsp12- T853R +8+7+RNA (185,245)	186,190 ± 50	Intact polymerase complex	22,300 ± 50	Excess nsp8 and/or RNA
SARS-CoV-2 nsp12 +8+8L7+RNA (185,603)	186,270 ± 40	Intact polymerase complex	110,600 ± 40	Free nsp12

Table 1: Native mass spectrometry of coronavirus polymerase complexes. Major and minor species from native mass spectrometry experiments are listed for each complex tested with their respective (\pm) standard deviations. The complexes best explained by the determined masses are listed next to each mass. Intact polymerase complexes are composed of one nsp7, two nsp8s, one nsp12, and one RNA duplex. Individual protein or RNA molecular weight are as follows for PEDV: 21,962 Da – RNA, 10,076 Da – nsp7, 21,688 Da – nsp8, 108,879 Da – nsp12, 31,384 Da – nsp8L7. Molecular weights for SARS-CoV-2 are as follows: 21,962 Da – RNA, 9,296 Da – nsp7, 21,938 Da – nsp8, 110,111 Da – nsp12, 31,592 Da – nsp8L7. (See also **Table S2** and **Figure S4**). The average and standard deviation for each complex were calculated from assigned spectral peaks in observed charge state distributions, where each spectrum is an average of 50 individual spectra.

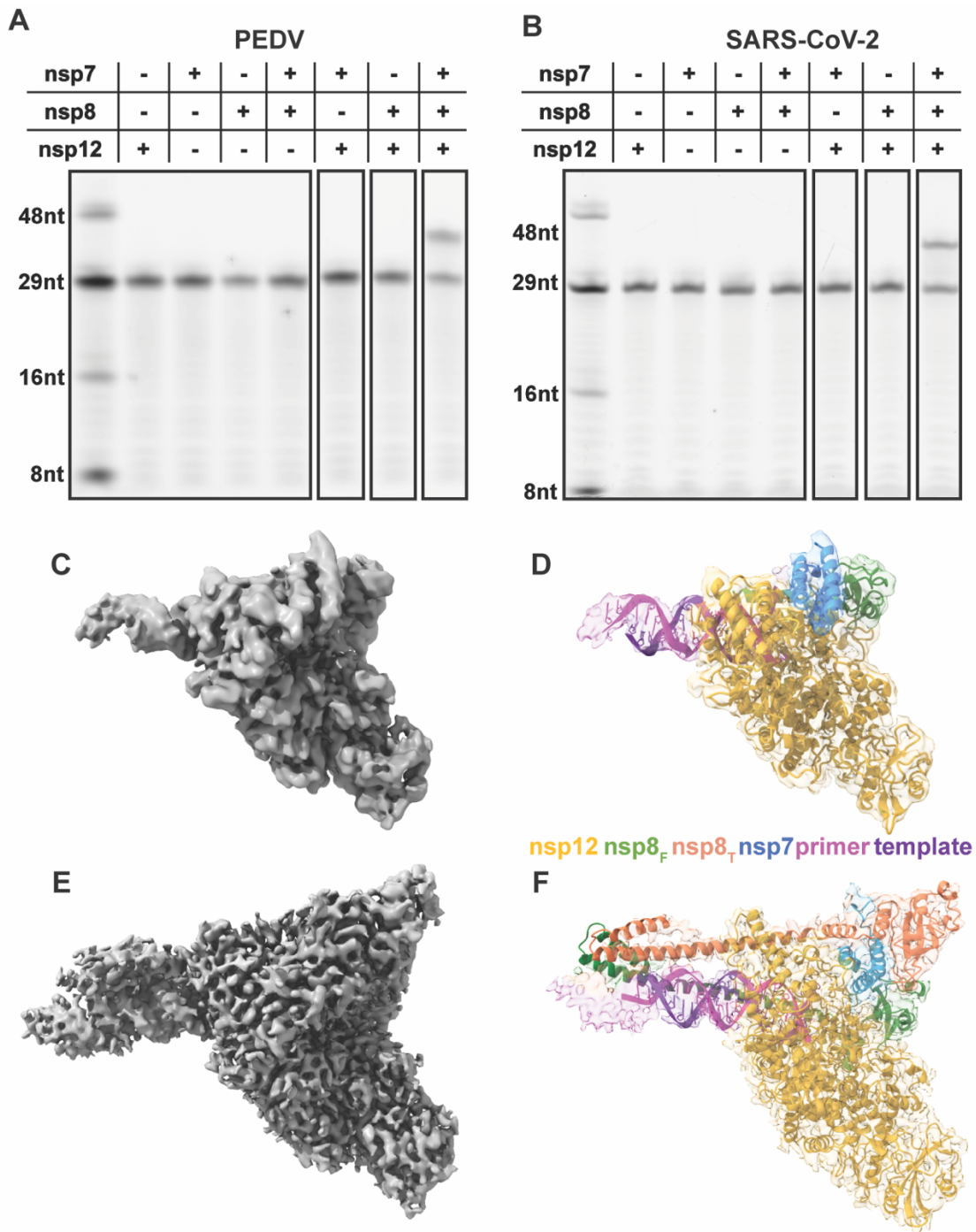


Figure 1: Assembly of an active PEDV polymerase complex. A 29 nt RNA primer with a 5' fluorophore is annealed to a 38 nt template and extended in the presence of CoV polymerase complexes. Combinations of nsp7, nsp8, and nsp12 were tested for PEDV (**A**) and SARS-CoV-2 (**B**). **C,E** 3.3 and 3.4 Å cryo-EM reconstruction of the PEDV core polymerase complex with (**E**) and without (**C**) nsp8_T, respectively. **D,F** Coordinate models of the PEDV core polymerase complexes docked into their corresponding electron density maps colored by chain.

Structure of PEDV core polymerase complex

Our initial attempts at cryo-EM structure determination using a short RNA duplex and DDM yielded a model of the PEDV core polymerase complex (nsp7, nsp8, nsp12) bound to a short RNA primer-template pair in a post-translocated state (**Fig. 1C and D**). Our final map has a resolution of 3.3 Å (**Table S1, Fig S1**). This cryo-EM map contains density for most of nsp12, including the NiRAN and polymerase domains. We resolved most of nsp7 (residues 2-62), the C-terminal portion of nsp8_F (residues 79-192), and a full turn of dsRNA exiting the polymerase active site. We lack well-resolved density for the nsp8_F N-terminal helical extension. 3D variability analysis showed that the dsRNA leaving the active site has flexibility, likely leading to the lack of density in the final reconstruction (**Video S1**) (36). Unexpectedly, this reconstruction lacked any density for nsp8_T, differing with most RNA-bound CoV polymerase structures as well as the PEDV polymerase native mass spectrometry data described above. To obtain the structure of the intact complex observed by native mass spectrometry we screened different detergents and grid preparation strategies. We found that the combined addition of the detergent CHAPSO and a longer RNA substrate promoted complex stability and allowed the determination of the complete PEDV core polymerase complex structure. Our more complete map has a resolution of 3.4 Å and lacks orientation bias. Importantly, in addition to previously resolved protein and RNA regions, the improved map has density for the N-terminal extension of nsp8_F and most of nsp8_T which was previously absent entirely.

Comparison of PEDV and SARS-CoV-2 polymerase core complex models

PEDV and SARS-CoV-2 nsp7, nsp8, and nsp12 protein sequences have significant sequence identity (nsp12 58.6%, nsp8 43.1%, and nsp7 41.8%) (**Fig. S6**). The

overall architecture of the PEDV polymerase core complex is very similar to published SARS-CoV-2 models with a nsp12 RMSD value of 0.964 Å (**Fig. 2A, S7**) (28). Comparison of our PEDV models' RdRP and NiRAN active sites with SARS-CoV-2 structures revealed highly conserved active site structures (**Fig. S7**). This structural conservation suggests that antivirals targeting SARS-CoV-2, such as Remdesivir or the recently studied dual-purpose AT-527, could be effective against PEDV and other alphaCoVs (41-43).

A major difference in the PEDV polymerase models is an altered loop conformation (PEDV nsp12 residues 249-268) that binds the nsp8_F head domain in PEDV but not in SARS-CoV-2 complexes (**Fig. 2B**). This conformational difference increases nsp8_F's buried surface area by 115 Å² in the PEDV polymerase core complex. This PEDV nsp12 loop sequence is well conserved within *AlphaCoV* but not across CoV genera (**Fig. S6**). Based on this conservation, we predict that the observed PEDV nsp12 loop conformation to interact with nsp8_F is shared among and specific to alphaCoVs.

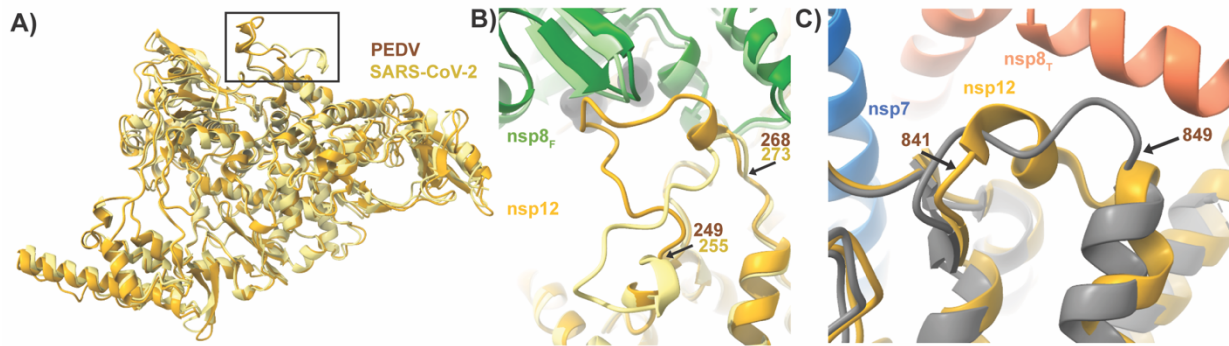


Figure 2: Structural differences between PEDV and SARS-CoV-2 nsp12s. **A)** Overlap of SARS-CoV-2 (PDB: 7KRP) and PEDV (8URB) nsp12. The black box highlights the loop region depicted in **(B)**. **B)** PEDV nsp12 loop (residues 249-268) buries extra surface area on nsp8_F compared to SARS-CoV-2 (PDB: 7KRP). **C)** PEDV nsp12 841-849 lacks helical definition in the absence of nsp8_T, while it has a short helix when the nsp8_T N-terminal extension is present. For each panel except **(C)** PEDV and SARS-CoV-2 models are superimposed with PEDV being colored in darker shades. In **(C)** nsp12 from our complex without nsp8_T is colored in grey. Superimposition of models was achieved using the Matchmaker function in ChimeraX (40).

A small difference between the two PEDV models is that in the model without nsp8_T the PEDV nsp12 residues 841-849 lack helical definition that is present in the complete core complex reconstruction (**Fig. 2C**). Helical definition in this region is also observed in SARS-CoV-2 models with resolved nsp8_T N-terminal helical extensions (PDB IDs: 6YYT, 7KRN, 7KRO, 6XEZ), but not in SARS-CoV (PDB ID: 6NUR) and SARS-CoV-2 nsp12 structures (PDB IDs: 7CYQ, 7CXM) without well resolved nsp8_T extensions region suggesting that association of the nsp8_T N-terminal helical region promotes helical secondary structure in this nsp12 region.

A PEDV nsp12 structure model bound to RNA without nsp8_T

Missing nsp8_T density in our initial reconstructed map was unexpected. As previously mentioned, there is a cryoEM structure of a betaCoV polymerase complex without nsp8_T that could bind RNA, so while our missing nsp8_T was surprising, it is not unprecedented (25). Prior structural work for betaCoV polymerases has proposed that both nps8_T and nsp8_F participate in CoV RNA synthesis, acting as “sliding-poles” that guide dsRNA out of the polymerase active site (23). Our native mass spectrometry results also contrast this PEDV polymerase structural model’s stoichiometry (**Table 1**). These observations led us to hypothesize that nsp8_T’s interactions with the polymerase complex are weaker than other replication factors (nsp8_F and nsp7), and perhaps the vitrification process during grid preparation caused nsp8_T disassociation. Despite the lability of nsp8_T during grid freezing, the PEDV nsp12 polymerase remained bound to nsp7, nsp8_F, and primer/template RNA in contravention to proposed models for polymerase complex assembly and activity (23). We therefore hypothesized that nsp8_T may not be required for CoV polymerase RNA binding and/or RNA synthesis.

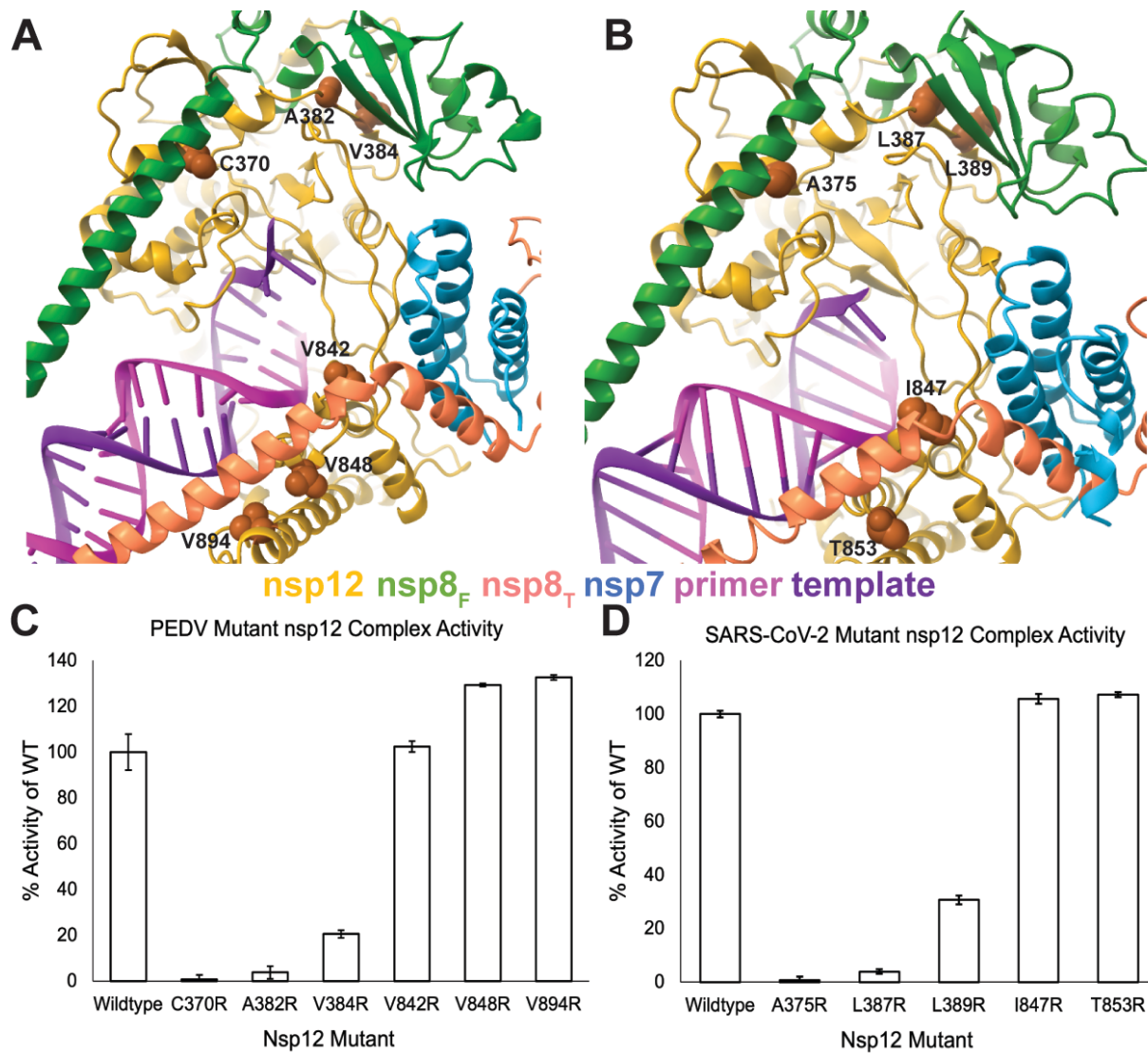


Figure 3: Effects of replication factor disrupting nsp12 mutations on polymerase activity. Sites of nsp12 mutations designed to disrupt replication factor interactions for **(A)** PEDV and **(B)** SARS-CoV-2 are shown as orange spheres. The effects of PEDV **(C)** and SARS-CoV-2 **(D)** nsp12 mutations were evaluated using *in vitro* RNA primer extension assays. Technical triplicates of reactions were run and percent activity of wildtype nsp12 is presented. Error bars indicate standard deviation of the triplicates. All reactions, except PEDV nsp12-V842R, were significantly ($P < 0.05$) different compared to the wildtype control. Significance was determined using an unpaired t-test for each mutant complex.

Prior studies have shown that MERS-CoV and SARS-CoV-2 nsp12 and nsp8 complexes can extend RNA *in vitro* (44,45). As the association of nsp8_T with the core complex is largely through binding to nsp7, even modest polymerase activity in the absence of nsp7 suggests a strong role for nsp8_F in stimulating polymerase activity, and the dispensability of nsp8_T for *in vitro* RNA synthesis. Under our reaction conditions, PEDV and SARS-CoV-2 polymerases require all three proteins (nsp12, nsp7, nsp8) for *in vitro* polymerase activity (**Fig. 1A and B**). To further dissect nsp8_F and nsp8_T's contributions to promoting polymerase activity, we designed homologous PEDV and SARS-CoV-2 nsp12 single amino acid mutations at each nsp8's protein-protein interfaces. These mutations were designed to be highly disruptive to hydrophobic protein-protein interfaces through the substitution of nsp12 surface residues with arginine to block interactions with regions of the nsp8s (PEDV nsp12 A382R, V384R, V842R, V848R and V894R and SARS-CoV-2 nsp12 A375R, L387R, L389R, I847T and T853R) (**Fig. 3A, B, and Fig S8**).

Native mass spectrometry (**Table 1, Fig. S4**) of select mutant complexes showed full complex stoichiometry, 1:2:1:1 nsp7:nsp8:nsp12:RNA (~186 kDa mass), for all mutants tested. The only exception being PEDV A382R where there were two major species: nsp12 alone and intact core polymerase complexes, indicating that nsp12 was either completely free or completely bound by nsp7 and nsp8 replication factors. These data suggest that the disruptive mutations are only effective in blocking the nsp12-nsp8 subdomain interactions and are not sufficient to fully prevent nsp8 subunit association with nsp12. This is congruent with the extended nature of nsp8 and the presence of both an N-terminal helical extension and a C-terminal head domain that bind distinct sites in

coronavirus core polymerase complexes for both nsp8_F and nsp8_T. Hence nsp12 mutations disrupting the binding of one region of nsp8 may not preclude association of that nsp8 with the core complex. The effects of these mutations on polymerase activity were examined using an *in vitro* primer extension assay (**Fig. 3C, D and Fig. S9**).

Nsp12 mutants designed to disrupt the association of PEDV nsp8_F N-terminal helical extension, (nsp12 C370R) or nsp8_F C-terminal head domain (nsp12 A382R or V384R) both resulted in complexes that either completely or mostly lost RNA extension activity (**Fig. 3C and Fig. S9**). In contrast, mutant complexes designed to disrupt association of nsp8_T N-terminal helical extension (nsp12 V842R, V848R, or V849R) retained RNA synthesis activity (**Fig. 3C**). These results indicate the importance of nsp8_F for PEDV RNA synthesis and demonstrate that association of the nsp8_T N-terminal helical extension with nsp12 is not required for stimulation of RNA synthesis *in vitro*.

To determine if these findings are specific to the PEDV polymerase core complex or a shared feature among CoVs, we tested homologous mutations for SARS-CoV-2 nsp12. Like PEDV, SARS-CoV-2 nsp12 mutants designed to disrupt association of the nsp8_F N-terminal extension (nsp12 A375R) or nsp8_F C-terminal head domain (nsp12 L387R or L389R) resulted in inactive polymerase complexes (**Fig. 3D and Fig. S9**). However, nsp12 mutations designed to disrupt association of the nsp8_T N-terminal extension (I847R or T853R) produced polymerase complexes with activity resembling that of the wild-type protein (**Fig. 3D**). Similar results for mutagenesis across both PEDV and SARS-CoV-2 core polymerase complexes suggest shared roles for nsp8s across the coronavirus subfamily. These data suggest that both the nsp8_F N-terminal extension and the C-terminal head domain play essential roles in stimulating polymerase activity beyond

simply facilitating protein-protein interactions while the nsp12 interaction with nsp8_T N-terminal extension appears non-essential for stimulating RNA synthesis.

Nsp8_T N-terminal RNA binding domain is not required to stimulate RNA synthesis activity

Recent work showed that a nsp8-nsp7 fusion protein (nsp8L7) can efficiently function as a nsp12 replication factor and allows the contributions of each nsp8 of a core complex's polymerase activity to be delineated (32). We produced SARS-CoV-2 and PEDV nsp8L7 fusion proteins with six-residue linker regions (**Fig. 4A, Fig. S3**). Both SARS-CoV-2 and PEDV nsp8L7 fusion proteins could stimulate their respective polymerase core complex activities in the presence of free nsp8 (**Fig. 4B, C and Fig. S5**). While SARS-CoV-2 nsp8L7 stimulated polymerase activity similar to wildtype nsp7 and nsp8, PEDV nsp8L7 only resulted in 20% stimulation compared to wildtype subunits. Since PEDV nsp12+nsp8 reactions have no RNA extension (figure **Fig. 4B and Fig. S5**) these data show that the PEDV nsp8L7 fusion protein can function as a cofactor, albeit with limited activity compared to wildtype proteins. We predict that the PEDV nsp8L7 reduction in polymerase stimulation is due to a reduced dissociation of the nsp8L7 fusion protein from oligomeric states into forms available for binding nsp12 leading to defects in polymerase assembly (32). Therefore, only a fraction of the available nsp12 may form full complexes as compared to SARS-CoV-2 nsp8L7 complexes leading to the reduced overall observed activity. Differences in the oligomeric states of *Alpha*- and *Beta*CoV nsp7-nsp8 cofactor complexes have been previously identified, supporting our prediction that PEDV nsp8L7 behaves differently than SAR-CoV-2 nsp8L7 (30). The prediction of incomplete PEDV core complex formation with nsp8L7 is supported by native mass spectrometry where a significant amount of PEDV nsp12 alone is observed (**Table 1, Fig**

S4). In support of nsp8_F being required for RNA synthesis *in vitro* and in validation of the fusion protein strategy to delineate nsp8 contributions, PEDV and SARS-CoV-2 nsp8L7 fusion proteins are not sufficient to stimulate CoV RNA synthesis in the absence of isolated nsp8 (**Fig 4B and C**).

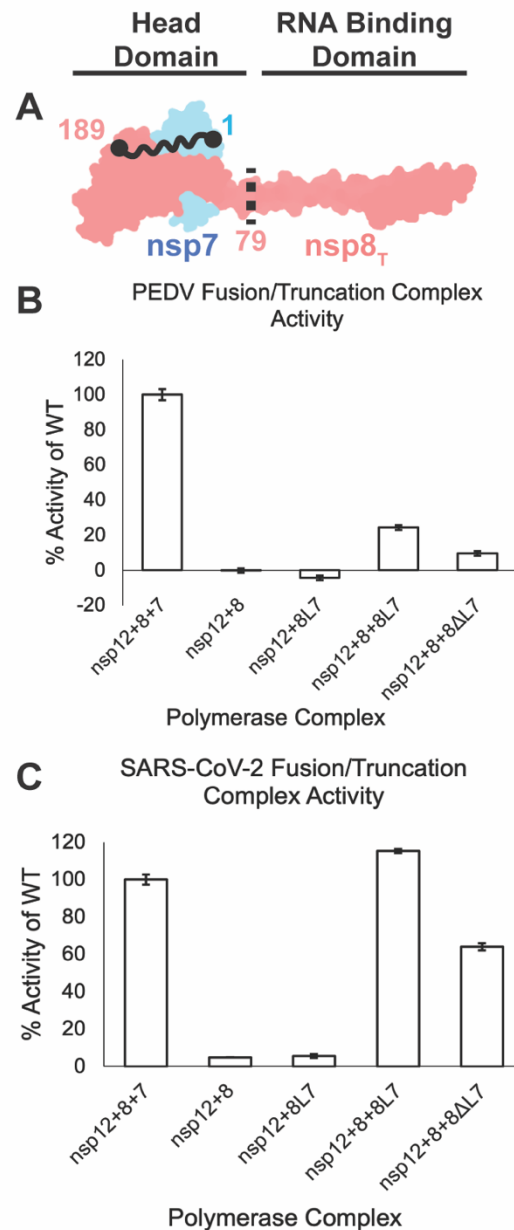


Figure 4: nsp8_T's N-terminal domain is not required for RNA synthesis. A) Model of nsp7 (blue) and nsp8_T (red) heterodimer (PDB ID: 7KRP). The black dots and squiggly line depict the nsp8 and nsp7 termini that are fused in nsp8L7. The dashed black line is the site of truncation of nsp8ΔL7, removing the nsp8_T N-terminal 79 amino acid RNA binding domain. **B)** PEDV and **C)** SARS-CoV-2 polymerase complex activity using the nsp8L7 and nsp8ΔL7 replication factors. Reactions were run in triplicate and percent activity was compared to a wildtype nsp7+nsp8+nsp12 complex. Error bars indicate standard deviation of the triplicates. Each reaction was determined to be significantly different than all other complexes tested ($p < 0.05$) using an unpaired t-test.

Using the PEDV and SARS-CoV-2 nsp8L7 fusion constructs, we produced 79 residue N-terminal truncations of nsp8_T (nsp8 Δ L7) which lack the previously described RNA binding region of the protein (**Fig. 4A, Fig. S3**) (23). Both PEDV and SARS-CoV-2 nsp8 Δ L7 stimulated polymerase primer extension activity *in vitro* (**Fig. 4 and Fig. S5**). Compared to polymerase stimulation using the nsp8L7 fusion constructs, PEDV and SARS-CoV-2 nsp8 Δ L7 each had an approximately 50% decrease in activity from their respective nsp8L7 reactions. These results indicate that the nsp8_T N-terminal region is not required for stimulation of RNA synthesis *in vitro*. Our results confirm similar prior results for SARS-CoV-2, while expanding the conclusions to PEDV (32). The differences in PEDV and SARS-CoV-2 polymerase activities could be attributed to the replication factor proteins' unique interactions across the CoV genera (30). While the magnitude of the ability of PEDV and SARS-CoV-2 nsp8 Δ L7 truncations to stimulate polymerase activity vary, that both truncation proteins can stimulate some polymerase activity while reactions lacking nsp7 do not, indicates that nsp8 Δ L7 is a functional replication factor and that the N-terminal extension of nsp8_T is not required for stimulation of *Alpha-* or *BetaCoV* RNA synthesis *in vitro*.

Discussion:

We have established that the PEDV core polymerase complex assembles into an active polymerase complex using a similar replication factor stoichiometry as was previously seen in betaCoVs. The structure of the PEDV core polymerase complex reveals an overall similar architecture to betaCoV complexes with one large conformational difference in PEDV nsp12 residues 249-269 to interact with nsp8_F that we predict to be conserved among alphaCoVs. Additionally, we used biochemistry and

mutagenesis to show that for both PEDV and SARS-CoV-2, nsp8_F is required for *in vitro* RNA synthesis while the N-terminal extension of nsp8_T is not.

Since the observed structural differences in the PEDV and SARS-CoV-2 core polymerase complexes lie outside of the common antiviral drug design targets (the polymerase and NiRAN active sites), we predict that antivirals targeting these shared sites would be effective against both *Alpha*- and *Beta*CoV. For example, structures have shown that SARS-CoV-2 nsp12 residue S861 is likely important for the effectiveness of the nucleotide analogue Remdesivir (46). The proposed mechanism is that S861 sterically clashes with the 1'-cyano group of the antiviral, impairing RNA elongation. For PEDV, residue S856 is conserved in both sequence and space to S861 of SARS-CoV-2 (**Fig. S7**), indicating that Remdesivir could be used to treat *Alpha*CoV infections (47).

After the initial observation of a labile nsp8_T interaction within the RNA-bound PEDV polymerase complex, we sought to explore the contribution of the nsp8 N-terminal extensions in simulating RNA synthesis activity. *In vitro* analyses with mutant PEDV and SARS-CoV-2 nsp12s and truncated nsp8_T (nsp8 Δ L7) established that nsp8_T's N-terminal extension is not required to stimulate *in vitro* RNA synthesis. Similar *in vitro* work with truncated nsp8s has already established that the RNA binding domains of each nsp8 are not required for processive RNA synthesis for SARS-CoV-2, but rather play important roles in complex assembly and RNA binding (32). In some published studies, SARS-CoV-2 and MERS-CoV complexes of just nsp12 and nsp8 are sufficient to produce some polymerase activity *in vitro* (44,45). The absence of nsp7 from these reactions indicates that nsp8_T is not absolutely required for *in vitro* RNA synthesis, further supporting our conclusions. Recent structural work has proposed that the N-terminal extensions of nsp8_F

and nsp8_T act as sliding poles to promote RNA synthesis and processivity (23). The sliding pole hypothesis provided a mechanistic explanation for prior observations that the mutation of a positively charged residue (K58A) in SARS-CoV nsp8's N-terminal extension was lethal to virus replication *in vivo* and reduced RNA synthesis *in vitro* (21). Our work was able to build upon these observations, by delineating the contributions of the two nsp8s and further confirm nsp8_F's importance in RNA synthesis. We propose a revision to this model where the nsp8_T N-terminal extension is not essential to stimulate RNA synthesis but may instead have other important roles in RNA synthesis.

We hypothesize that nsp8_T may act as a scaffolding protein to regulate the engagement of other viral factors like the nsp13 RNA helicase to promote RNA backtracking (**Fig. 5**). Previous work has shown that the viral helicase, nsp13, binds template RNA entering the active site and can cause polymerase backtracking on primer-template RNA (27,28). In a backtracked polymerase state, the 3' end of the nascent RNA extrudes out of the polymerase NTP entry channel. While a function for backtracking within viral replication remains unclear, one hypothesis is that it mediates proofreading of mis-incorporated 3' nucleotides during CoV RNA synthesis. The nsp13 responsible for backtracking RNA binds the polymerase core complex via the nsp8_T N-terminal extension. As a potential model for viral RNA proofreading, we propose that aberrant nucleotide incorporation stalls elongation, allows engagement of the nsp8_T N-terminal extension with nsp12 and the RNA allowing the backtracking nsp13 RNA helicase to bind template RNA and induce a backtracked state. 3' nucleotides of the backtracked RNA could then be excised by the viral nsp14 exonuclease removing the mis-incorporated nucleotides.

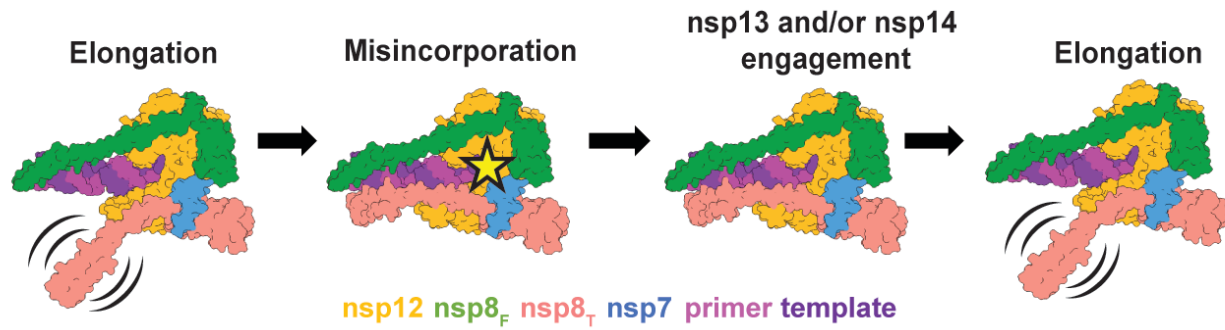


Figure 5: Model for coronavirus backtracking and proofreading. Stalling of polymerase complexes by misincorporation is hypothesized to promote the engagement of nsp8_T's N-terminal helical extension with RNA allowing the recruitment of additional viral factors for backtracking or proofreading. Whether or not nsp8_T is always associated with the complex is unknown.

Our current work is focused on the multiple roles of nsp8 in *Alpha*- and *Beta*CoV RNA synthesis. To date, there are no reports of active *Gamma*- or *Delta*CoV polymerase complexes *in vitro*, or published structures of polymerases from these two genera. Sequence conservation of the nsps involved in genome replication (i.e. nsp7, nsp8, and nsp12) across all CoV genera indicates that these proteins have conserved roles in replication (48). To our knowledge nsp7 does not directly interact with RNA during replication, but it serves multiple important roles to support CoV RNA synthesis. Beyond nsp7's role as a replication factor for nsp12 and scaffolding protein for nsp8_T, nsp7 also de-oligomerizes nsp8 allowing monomeric nsp8 to interact with nsp12 as nsp8_F (21,30,49). Furthermore, the interactions between nsp7 and nsp8 oligomers are known to vary across the CoV species (30). These differences are possible explanations for the different activities of nsp8_{L7} and nsp8_{ΔL7} containing complexes seen in our work.

This work expands the diversity of CoV polymerase structural biology to the alphaCoVs. Our work shows a high level of structural conservation among CoV replication machinery and has allowed for the generation of hypotheses for CoV RNA synthesis that span more than a single virus genus. Such work is essential not only for a greater understanding of CoV biology but also in preparing for the threat of emerging CoV.

Data Availability:

The PEDV core polymerase complex electron density maps have been deposited in the Electron Microscopy Data Bank for the complex lacking nsp8_T (EMDB: 29779) and the complete complex (EMDB: 42488). Coordinate models have been deposited in the Protein Data Bank for the complex lacking nsp8_T (PDB: 8G6R) and the complete complex (PDB: 8URB).

Limitations of Study:

We acknowledge that our study uses short RNA substrates, short extension reactions, and has been done exclusively *in vitro* and so does not model processive CoV RNA synthesis. This limits our ability to make conclusions on nsp8_T's importance for processive RNA synthesis, and replication *in vivo*. We believe studies using longer RNA substrates, and *in vivo* reverse genetics will allow our hypotheses and conclusions developed here to be evaluated in the context of processive replication and virus replication *in vivo* but are beyond the scope of the current work.

Our work only studies the core polymerase complex (nsp7, nsp8, and nsp12) but our conclusions include hypotheses about interactions with other viral proteins (i.e., nsp13 and nsp14). While our current work is unable to directly assess these hypotheses, this work provides a foundation for further study.

Funding:

This work was supported by NIH/NIAID AI123498 and AI158463 and USDA WIS03099 grants to R.N.K, as well as NIH R35GM118110 grant to J.J.C.

Acknowledgements:

We would like to thank the Cryo-EM Research Center in the Biochemistry Department at University of Wisconsin-Madison for technical and staff support. We would like to thank Dr. Olve Peersen for his suggestions and insight for use of the nsp8-nsp7 fusion construct.

Video S1 can be found at: <https://doi.org/10.1093/nar/gkae153>

References:

1. Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, **1282**, 1-23.
2. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.
3. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. and Fouchier, R.A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, **367**, 1814-1820.
4. de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R.A., Galiano, M., Gorbalenya, A.E., Memish, Z.A. *et al.* (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, **87**, 7790-7792.
5. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265-269.
6. Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L. *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*, **579**, 270-273.
7. Silva, C.S., Mullis, L.B., Pereira, O., Jr., Saif, L.J., Vlasova, A., Zhang, X., Owens, R.J., Paulson, D., Taylor, D., Haynes, L.M. and Azevedo, M.P. (2014) Human Respiratory Coronaviruses Detected In Patients with Influenza-Like Illness in Arkansas, USA. *Virology*, **2014**, 1-8.
8. Xiu, L., Binder, R.A., Alarja, N.A., Kochek, K., Coleman, K.K., Than, S.T., Bailey, E.S., Bui, V.N., Toh, T.H., Erdman, D.D. and Gray, G.C. (2020) A RT-PCR assay for the detection of coronaviruses from four genera. *J Clin Virol*, **128**, 104391.
9. Vlasova, A.N., Diaz, A., Dantie, D., Xiu, L., Toh, T.H., Lee, J.S., Saif, L.J. and Gray, G.C. (2022) Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis*, **74**, 446-454.
10. Yan, Q., Liu, X., Sun, Y., Zeng, W., Li, Y., Zhao, F., Wu, K., Fan, S., Zhao, M., Chen, J. and Yi, L. (2022) Swine Enteric Coronavirus: Diverse Pathogen-Host Interactions. *Int J Mol Sci*, **23**, 1-23.
11. Liu, Q. and Wang, H.Y. (2021) Porcine enteric coronaviruses: an updated overview of the pathogenesis, prevalence, and diagnosis. *Vet Res Commun*, **45**, 75-86.

12. Zhang, Y., Chen, Y., Zhou, J., Wang, X., Ma, L., Li, J., Yang, L., Yuan, H., Pang, D. and Ouyang, H. (2022) Porcine Epidemic Diarrhea Virus: An Updated Overview of Virus Epidemiology, Virulence Variation Patterns and Virus-Host Interactions. *Viruses*, **14**.
13. Stevenson, G.W., Hoang, H., Schwartz, K.J., Burrough, E.R., Sun, D., Madson, D., Cooper, V.L., Pillatzki, A., Gauger, P., Schmitt, B.J. *et al.* (2013) Emergence of Porcine epidemic diarrhea virus in the United States: clinical signs, lesions, and viral genomic sequences. *J Vet Diagn Invest*, **25**, 649-654.
14. Sun, R.Q., Cai, R.J., Chen, Y.Q., Liang, P.S., Chen, D.K. and Song, C.X. (2012) Outbreak of porcine epidemic diarrhea in suckling piglets, China. *Emerg Infect Dis*, **18**, 161-163.
15. Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.
16. Xu, X., Liu, Y., Weiss, S., Arnold, E., Sarafianos, S.G. and Ding, J. (2003) Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res*, **31**, 7117-7130.
17. te Velthuis, A.J., Arnold, J.J., Cameron, C.E., van den Worm, S.H. and Snijder, E.J. (2010) The RNA polymerase activity of SARS-coronavirus nsp12 is primer dependent. *Nucleic Acids Res*, **38**, 203-214.
18. Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res*, **96**, 59-126.
19. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
20. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2020) Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell*, **184**, 184-193.
21. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
22. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.T., Meng, X., Xia, H., Wood, J.M., Harris, L.D. *et al.* (2021) Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife*, **10**.

23. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
24. Yin, W., Mao, C., Luan, X., Shen, D.D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M. *et al.* (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science*, **368**, 1499-1504.
25. Bravo, J.P.K., Dangerfield, T.L., Taylor, D.W. and Johnson, K.A. (2021) Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Molecular Cell*, **81**, 1548-1552.
26. Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M. *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun*, **11**, 5874.
27. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T. *et al.* (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
28. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, **118**.
29. Xiao, Y., Ma, Q., Restle, T., Shang, W., Svergun, D.I., Ponnusamy, R., Sczakiel, G. and 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.
30. Krichel, B., Bylapudi, G., Schmidt, C., Blanchet, C., Schubert, R., Brings, L., Koehler, M., Zenobi, R., Svergun, D., Lorenzen, K. *et al.* (2021) Hallmarks of Alpha- and Betacoronavirus non-structural protein 7+8 complexes. *Sci Adv*, **7**.
31. Ma, M., Yang, Y., Wu, L., Zhou, L., Shi, Y., Han, J., Xu, Z. and Zhu, W. (2022) Conserved protein targets for developing pan-coronavirus drugs based on sequence and 3D structure similarity analyses. *Comput Biol Med*, **145**, 105455.
32. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.
33. Schneider, C.A., Rasband, W.S. and Eliceiri, K.W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*, **9**, 671-675.
34. Mastronarde, D.N. (2005) Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol*, **152**, 36-51.

35. Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, **14**, 290-296.
36. Punjani, A. and Fleet, D.J. (2021) 3D variability analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. *J Struct Biol*, **213**, 107702.
37. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.
38. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A. and Adams, P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol*, **74**, 531-544.
39. Croll, T.I. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol*, **74**, 519-530.
40. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.A.-O. and Ferrin, T.A.-O. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, **30**, 70-82.
41. Shannon, A., Fattorini, V., Sama, B., Selisko, B., Feracci, M., Falcou, C., Gauffre, P., El Kazzi, P., Delpal, A., Decroly, E. *et al.* (2022) A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. *Nat Commun*, **13**, 621.
42. Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., Smith, E.C., Case, J.B., Feng, J.Y., Jordan, R. *et al.* (2018) Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio*, **9**, e00221-00218.
43. Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P. and Gotte, M. (2020) Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem*, **295**, 6785-6797.
44. Gordon, C.J., Tchesnokov, E.P., Feng, J.Y., Porter, D.P. and 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.
45. Pourfarjam, Y., Ma, Z. and Kim, I.K. (2022) ATP enhances the error-prone ribonucleotide incorporation by the SARS-CoV-2 RNA polymerase. *Biochemical and Biophysical Research Communications*, **625**, 53-59.
46. Kokic, G., Hillen, H.S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Hobartner, C. and Cramer, P. (2021) Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun*, **12**, 279.

47. Krentz, D., Zenger, K., Alberer, M., Felten, S., Bergmann, M., Dorsch, R., Matiasek, K., Kolberg, L., Hofmann-Lehmann, R., Meli, M.L. *et al.* (2021) Curing Cats with Feline Infectious Peritonitis with an Oral Multi-Component Drug Containing GS-441524. *Viruses*, **13**.
48. Grellet, E., L'Hote, I., Goulet, A. and Imbert, I. (2022) Replication of the coronavirus genome: A paradox among positive-strand RNA viruses. *J Biol Chem*, **298**, 101923.
49. Biswal, M., Diggs, S., Xu, D., Khudaverdyan, N., Lu, J., Fang, J., Blaha, G., Hai, R. and Song, J. (2021) Two conserved oligomer interfaces of NSP7 and NSP8 underpin the dynamic assembly of SARS-CoV-2 RdRP. *Nucleic Acids Res*, **49**, 5956-5966.

Supplemental Figures and Data:

EMDB	29779	42488
PDB	8G6R	8URB
Microscope	Titan Krios	Talos Arctica
Voltage (kV)	300	200
Detector	K3 direct electron detector (Gatan)	K3 direct electron detector (Gatan)
Dose Rate (e ⁻ /pixel/sec)	12	14.1
Exposure Time (sec)	3.5	4.8
Electron Exposure (e ⁻ /Å ²)	60	60
Frames (no.)	60	60
Defocus Values	-0.7 to -2.5	-0.5 to -2.0
Sample Tilt (°)	25	0
Data Collection Mode	EFTEM, Counting, CDS	EFTEM, Counting, CDS
Nominal Magnification	105,000x	79,000x
Pixel Size (Å)	0.834	1.064
Symmetry Imposed	C1	C1
Movies Collected (no.)	1,261	5,806
Initial Particle Images (no.)	895,007	2,814,477
Final Particle Images (no.)	74,367	103,940
Map Resolution (Å) – GSFSC	3.3	3.4
Initial Models Used (PDB ID)	7CYQ	8G6R, 7KRP
Non-hydrogen Atoms	9,406	11,657
Protein Residues	1,075	1,359
Nucleic Acid Residues	40	41
Other Atoms	2 (Zn ²⁺)	2 (Zn ²⁺)
R.M.S. Deviations		
Bond Lengths (Å)	0.004	0.003
Bond angles (°)	0.556	0.887
MolProbity Score	1.64	1.13
Clashscore	7.19	2.86
Ramachandran Plot		
Favored (%)	96.34	97.78
Allowed (%)	3.66	2.22
Disallowed (%)	0	0

Table S1: Cryo-EM data collection and refinement. Information provided is for the cryo-EM data collection, and processing that produced the electron density map of the PEDV polymerase complexes. Additionally, validation statistics for the polymerase complex coordinate models built into the reconstructions are provided.

Sample	Desolvation Parameters	
	Capillary Temperature (°C)	In-source Trapping (V)
PEDV nsp12+8+7	250	-150
PEDV nsp12+8	250	-300
PEDV A382R-nsp12+8+7	200	-300
PEDV V848R-nsp12+8+7	200	-200
PEDV nsp12+8+8L7	250	-300
SARS-CoV-2 nsp12+8+7	200	-150
SARS-CoV-2 nsp12+8	200	-300
SARS-CoV-2 L387R-nsp12+8+7	200	-150
SARS-CoV-2 T853R-nsp12+8+7	200	-160
SARS-CoV-2 nsp12+8+8L7	250	-150

Table S2, Native mass spectrometry desolvation parameters: Parameters listed were optimized for each sample to remove as many adducts as possible without disrupting the complex(es) of interest.

Supplemental video 1, 3D variability analysis of PEDV core complex lacking nsp8_T:

The movie displays a series of 3D volumes (in gray) determined by 3D variability analysis (cryoSPARC v3.3.1). The movie begins by oscillating between the extremes of variability within the volume series. Docking our PEDV complex lacking nsp8_T model (nsp12 – light purple, nsp7 – green, nsp8 – red, primer – neon green, template – dark purple) into the densities reveals the flexibility of dsRNA leaving the active site. We predict that this movement results in the lack of complete reconstruction of our dsRNA and nsp8_F N-terminal extension in our final map and model of the PEDV complex lacking nsp8_T.

(<https://doi.org/10.1093/nar/gkae153>)

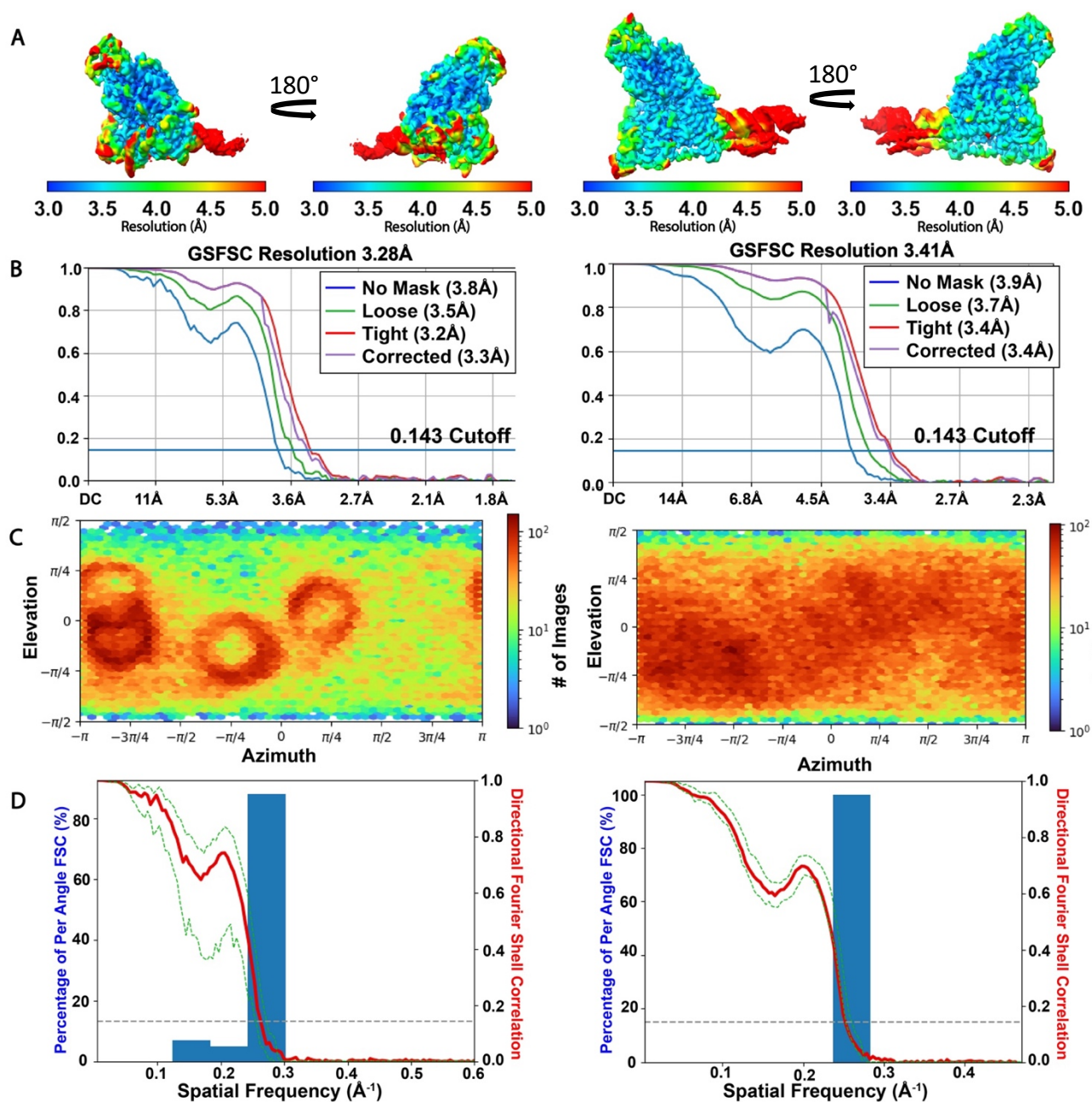
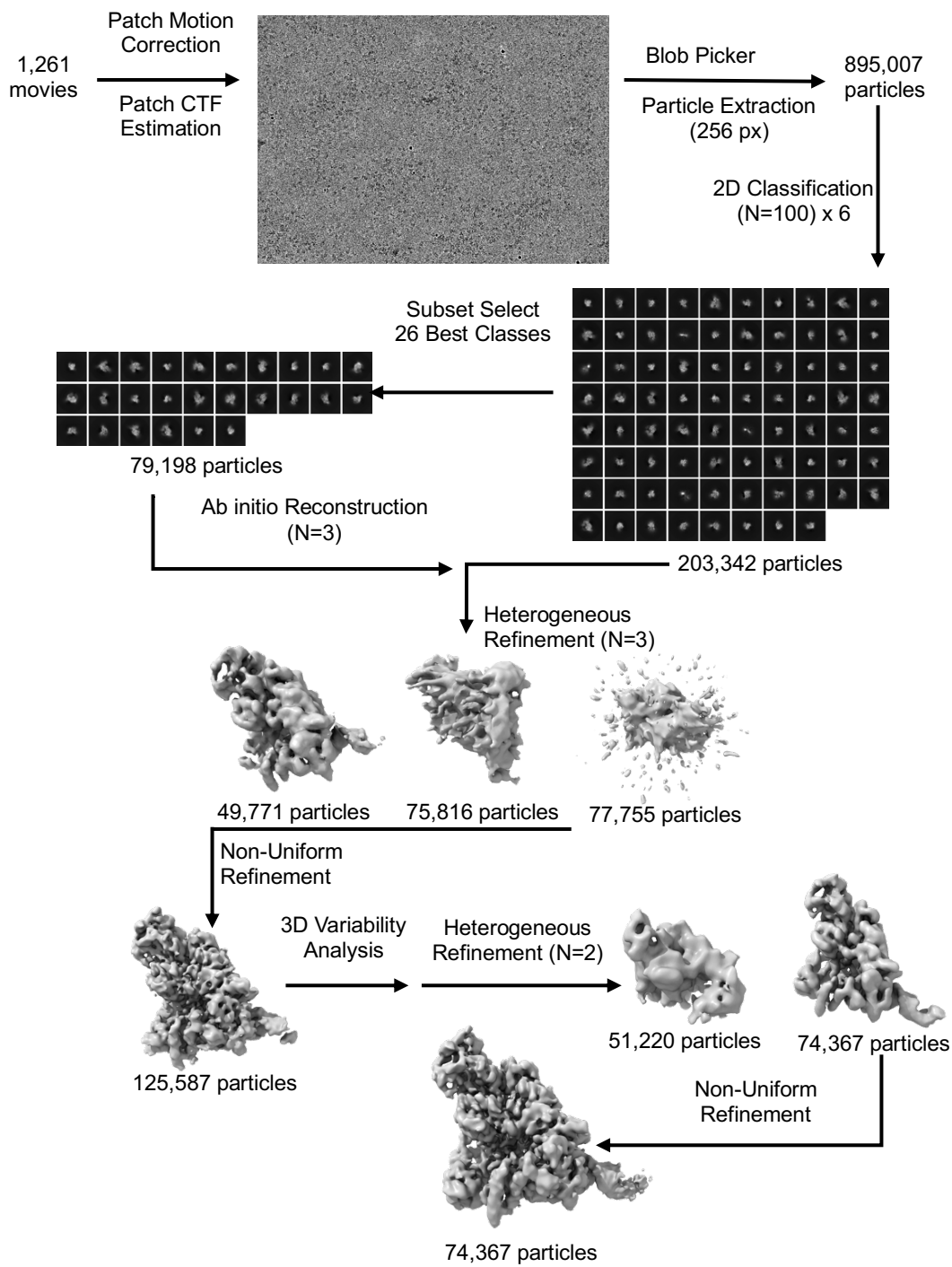


Figure S1, Cryo-EM validation: Data for the reconstruction missing nsp8_T (EMDB 29779) is on the left, and data for the complete complex reconstruction (EMDB 42488) is on the right. **A**) Cryo-EM reconstruction colored by local resolution according to the color key. **B**) Gold-standard Fourier shell correlation (FSC) plot determined in cryoSPARC. The solid blue line marks the 0.143 cutoff. **C**) Particle orientation distribution plot for the final reconstruction. “# of images” indicates number of projections at a particular orientation (Elevation x Azimuth). **D**) 3D FSC histogram plot for the final reconstruction. For 29779

the calculated sphericity was 0.818, with a global resolution 3.81 Å (unmasked). For 42488 the calculated sphericity was 0.864, with a global resolution of 4.01 Å (unmasked). The red line is the global FSC, the green lines are ± 1 SD from the mean of the directional FSC, and the blue histogram is the directional FSC. 3D FSC was calculated as described (1).

A)



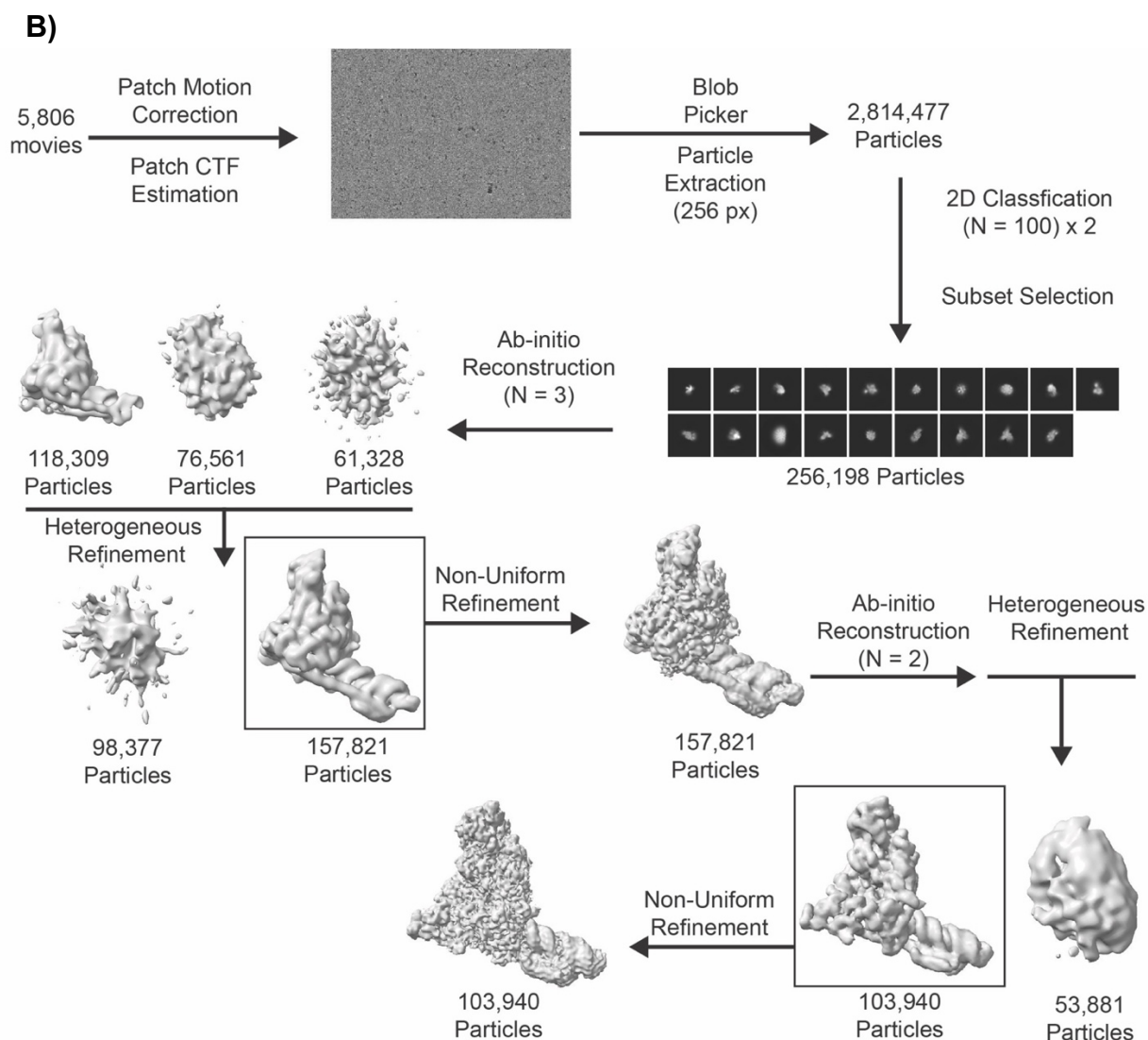


Figure S2, Cryo-EM processing pipeline: Workflows for the processing of cryo-EM data of **A)** 29779 (lacking nsp8_T) and **B)** 42488 (full model). All data processing was performed in cryoSPARC (2). Poor particles were removed through multiple rounds of 2D classification, then a subset of polymerase classes was used to generate initial models for subsequent heterogeneous refinement with all remaining particles. Particles from classes resembling polymerases were merged for further 3D refinement and a final round of heterogeneous refinement into polymerase complex maps. The complex maps were further refined to our final 3.3 or 3.4 Å reconstructions for complexes without or with nsp8_T respectively.

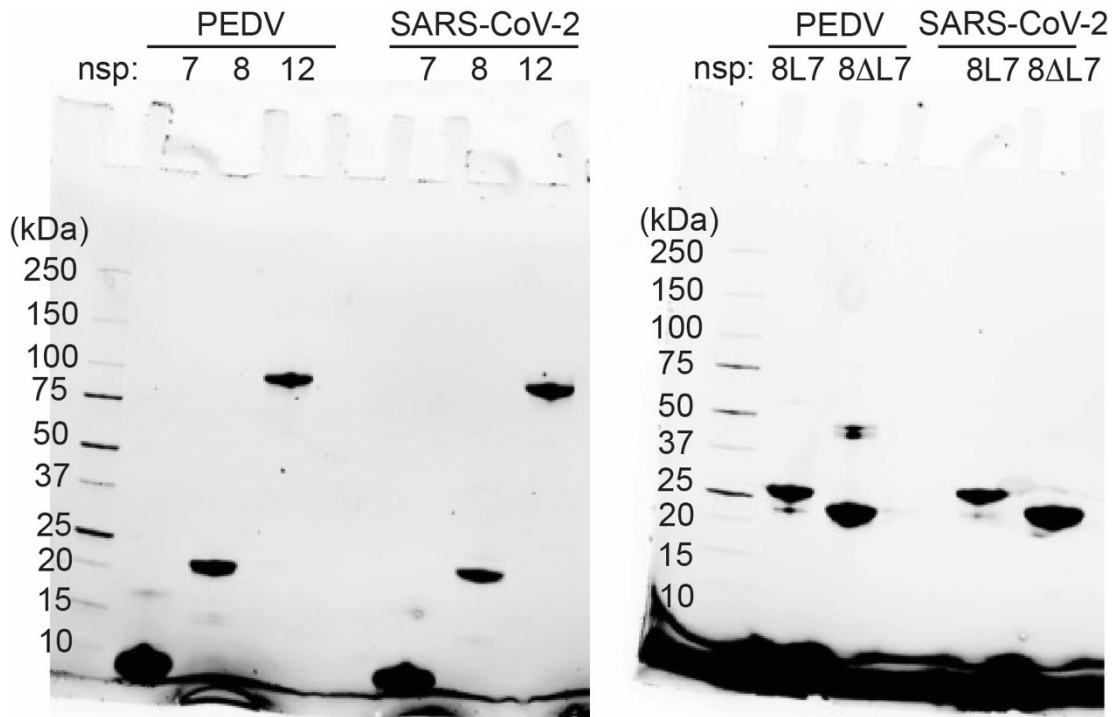
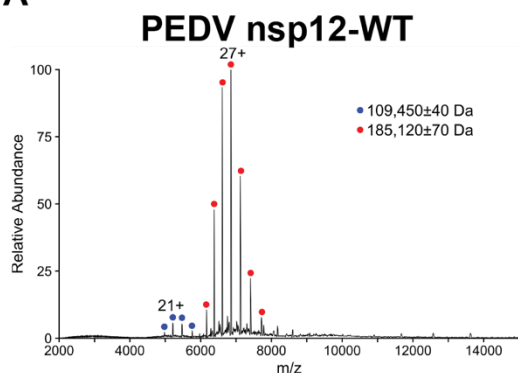
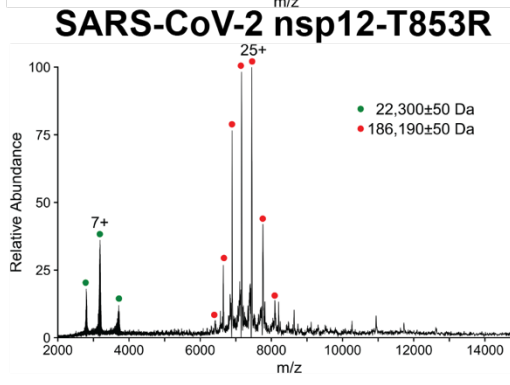
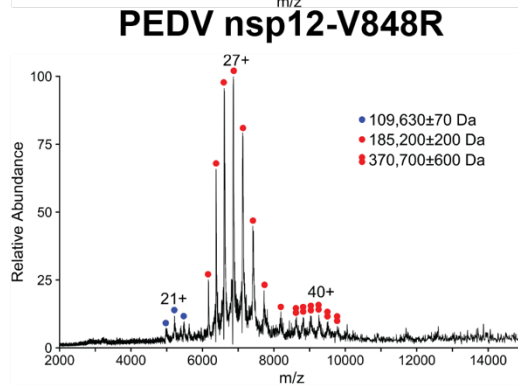
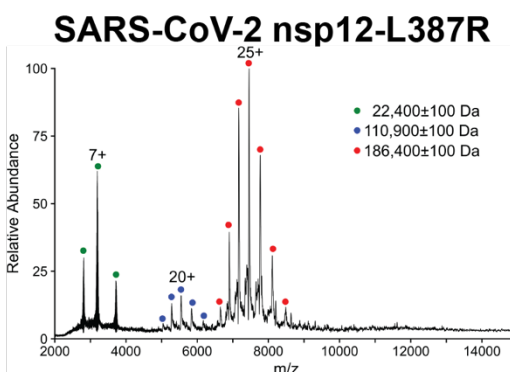
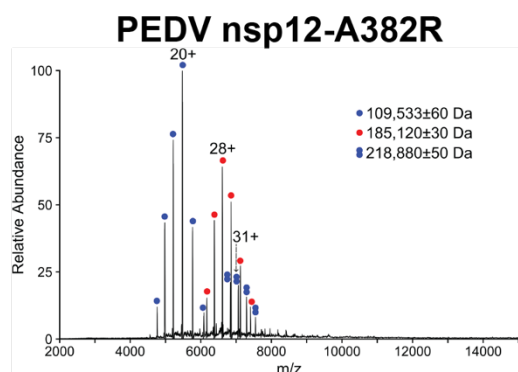
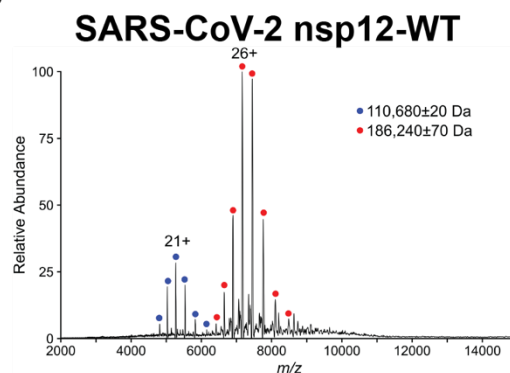


Figure S3, SDS-PAGE of viral proteins: Samples of purified proteins used for *in vitro* studies were run on pre-cast, stain-free, 4-20% SDS-PAGE gels (BioRad) and imaged using UV fluorescence. Expected molecular weights for different proteins are nsp7 – 9 kDa, nsp8 – 22 kDa, nsp12 – 110 kDa, nsp8L7 – 31 kDa, and nsp8ΔL7 – 25 kDa. The far-left lane of each gel is a protein ladder protein MWs labeled.

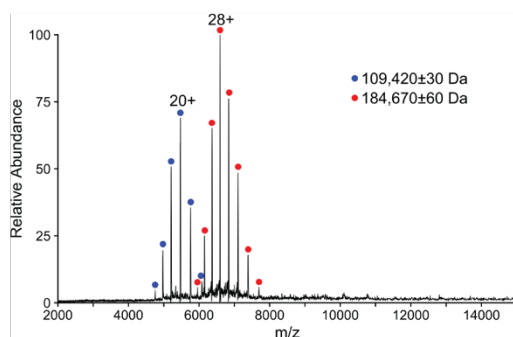
A



B



PEDV nsp12+nsp8+nsp8L7



SARS-CoV-2 nsp12+nsp8+nsp8L7

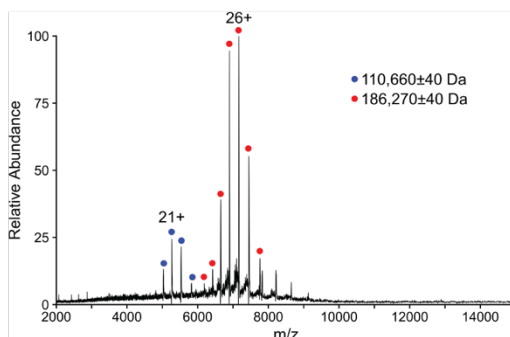


Figure S4, Native mass spectrometry of coronavirus polymerase complexes. Native mass spectra for all coronavirus polymerase complexes tested with major mass populations labelled for each. Single and double red dots are monomeric and dimeric full intact complexes, respectively. Single and double blue dots are monomeric and dimeric solo nsp12, respectively. Green dots are free nsp8. **A)** SARS-CoV-2 complexes, top to bottom: wildtype, nsp12+8, L387R, T853R, nsp8L7. **B)** PEDV complexes, top to bottom: wildtype, nsp12+8, A382R, V848R, nsp8L7. The most intense peak for each group has its charge state labelled above the peak. The average and standard deviation for each complex were calculated from 50 individual spectra.

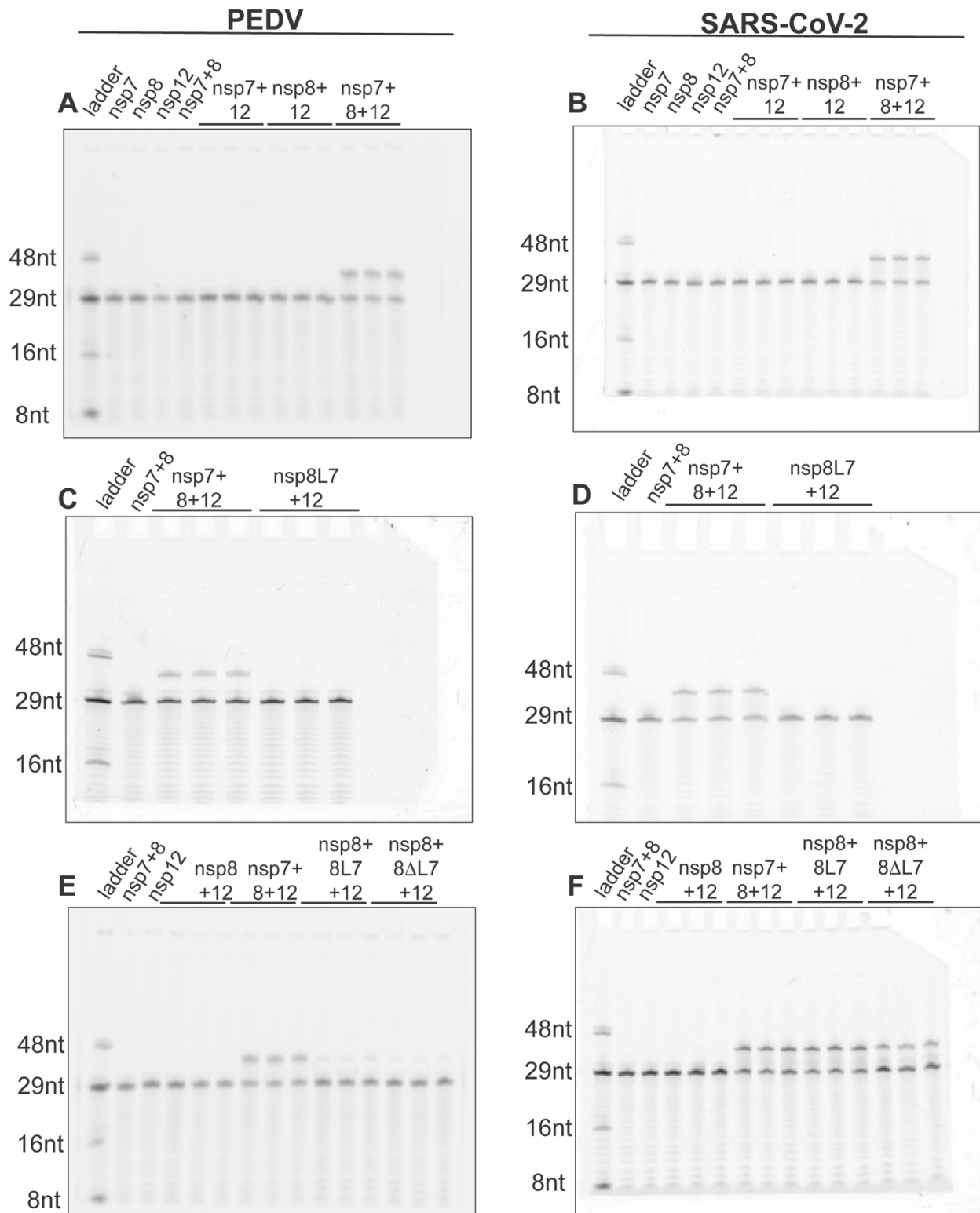


Figure S5, Full gel images for piecewise and fusion protein primer extensions: Primer extensions assess the activity of nsp12 by its ability to extend an RNA primer (29 nucleotides or nt) to the length of its RNA template (38 nt) pair. **A)** PEDV piecewise primer

extension with nsp12 and its replication factors nsp7 and nsp8. **B)** SARS-CoV-2 piecewise primer extension. Primer extension testing the nsp8L7 fusion protein's ability to stimulate nsp12 activity for PEDV (**C**) and SARS-CoV-2 (**D**). Primer extension with the nps8L7 and nsp8 Δ L7 fusion proteins and free nsp8 testing these replication factor's ability to stimulate nsp12 activity for PEDV (**E**) and SARS-CoV-2 (**F**).

δ	HKU19	-LGYSSNQNNSYLNRVKGSS-DARLEPCTSDNRPDVVRVAFNIYNN--ATAGIFKSTKNN	56
	HKU11	-----NSPYLNRVTGSS-GARLEPQQPGVTPDAVKRAFVHNN--TTSIGIFLSTKTN	49
α	PDCV	-----NSAYLNRVTGSS-DARLEPLQPGTQPDVAVKRAFVHND--TTSIGIFLSTKSN	49
	FCoV_65F	----GTTVDQSYLNRVGRSS-AARLEPCN-GTDPDHVSRAFDIYNK--DVACIGKFLKTN	52
α	HCoV_229E	-----SFDSSYLNRVGRSS-AARLEPCN-GTDIDYCVRAFVDVYNK--DASFIGKNLKSNN	50
	HCoV_NL63	-----SVDISYLNRRARGSS-AARLEPCN-GTDIDKCVRAFDIYNK--NVSFLGKCLKMN	50
γ	HKU8	-----SLDNNYLNRVGRSS-AARLEPCN-GTEPEHVIRAFDIYNK--DVACIGKFKVKN	50
	PEDV	-----STDMAYLNRVGRSS-AARLEXCN-GTDTQHVYRAFDIYNK--DVACLKGFLLKVN	50
γ	HKU22	-----SVPQNYLNRVGRGLS-EARLRPCASGLLPDVVKRAFIDLNS--NTAGMYASLKHNN	51
	IBV	SVAVASGFDKNYLNRVGRSS-EARLIPLANGCDPDVVKRAFVDCNK--ESAGMFQNLKRN	57
β	HCoV_OC43	-----SKDTNFLNRVGRGASVDARLVPCASGLSTDVQLRAFDIYNA--SVAGIGLHLKVN	52
	HCoV_HKU1	-----SKDLNFLNRVGRGTSVNARLVPCASGLSTDVQLRAFIDICNT--NRAGIGLYYKVN	52
β	MHV	-----SKDTNFLNRVGRGTSVNARLVPCASGLSTDVQLRAFIDICNA--NRAGIGLYYKVN	52
	MERS	-----SKDSNFLNRVGRGTSVNARIEPCSSGLSTDVVFRAFIDICNYKAKVAGIGKYKTN	54
β	HKU4	-----SKDTNFLNRVGRGSSVNARLEPCSSGLTDDVYRAFIDICNFARVAGIGKYKTN	54
	HKU5	-----SKDSNFLNRVGRGTSVNARIEPCASGLTDDVVFRAFIDICNYKAKVAGIGKYKTN	54
β	HKU9	-----AKDECFLNRVGRGTSVVARLVPLGSGVQPDIVLRAFIDICNT--KVAGFGLHLKNN	52
	SARS-CoV-2	-----SADAQSFNLRVCGVSS-AARLTPCGTGTSTDVYRAFDIYND--KVAGFAKFLKTN	52
β	SARS-CoV	-----SADASTFLNRVCGVSS-AARLTPCGTGTSTDVYRAFDIYNE--KVAGFAKFLKTN	52

:***. * ** : . : **.: * : : *

δ	HKU19	CTRFKSTRPGSI INKPVRTIETFFVTKQCTENVFRAEEQCYNMLPKSIVSTDDKFCVAV	116
	HKU11	CSRFRITTKQNLPLP-NKGSVELYFVSKQCSQQVFEIEETCYNMFDDSLKSTPEKFGVLAR	108
α	PDCV	CARFKTTRSALPLP-NKGEVELYFVTKQCAKVFIEIEECYNALSTELYTTDDTFGVLAK	108
	FCoV_65F	CSRFRNLD-----KRDAYYVVKRCKTSVMDHEQVCYNDLKD-----SGAVAE	94
α	HCoV_229E	CVRFKNVD-----KDDAFYIVKRCIKSVMDHEQSMYNLLKG-----CNAVAK	92
	HCoV_NL63	CVRFKNAD-----LKDGYFVVKRCKTSVMEHEQSMYNLLNF-----SGALAE	92
γ	HKU8	CVRFKNAD-----KHDAFYVVKRCKTSVMEHEQSIYDALKD-----CGAVSP	92
	PEDV	CVRLKNLD-----KHDAFYVVKRCKTSAMEHEQSIYSRLEK-----CGAVAE	92
γ	HKU22	CARFQELDENDD-----EIDSEFFVVKQTPPHNFEHEEKCYLDLKA-----DCVAV	96
	IBV	CARFQEVDRDTEG-----GNLEYCDSYFVVKQTPPSNYEHEKACYEDLKS-----EVTAD	106
β	HCoV_OC43	CCRFRQVDENGD-----KLDQFFVVKRDTLTIYNREMKCYERVKD-----CKFVAE	98
	HCoV_HKU1	CCRFRQVDENGD-----KLDKFFVVKRDTNLEVYNKEKTYELTKS-----CGVVAE	98
β	MHV	CCRFRQVDENGD-----KLDKFFVVKRDTNLEVYNKEKTYELTKS-----CGVVAE	98
	MERS	TCRFVELDDQGH-----HLDSYFVVKRHTMENYELEKHCYDILLRD-----CDAVAP	100
β	HKU4	TCRFVQVDDQGH-----KLDSEYFVVKRHTMENYELEKRCYDILLKD-----CDAVAI	100
	HKU5	TCRFVEVDDQGH-----RLDSYFVVKRHTMENYELEKRCYDVLKD-----CDAVAV	100
β	HKU9	CCRYQELDADGN-----QLDSYFVVKRHTESNYLLEQRCYEKLD-----CDVVAR	98
	SARS-CoV-2	CCRFRQEKDEDDN-----LIDSEYFVVKRHTFSNYQHEETIYNLLKD-----CPAVAK	98
β	SARS-CoV	CCRFRQEKDEEGN-----LLDSYFVVKRHTMSNYQHEETIYNLVKD-----CPAVAV	98

NiRAN Motif A_n

* : :: * : * * :

δ	HKU19	HDFFKFDG---VNNVVRHLTKYTLDDLVAALRHLSTS-QEIIQEILITMCGTSEE---	168
	HKU11	TEFFKFDD---IPNVNRQFLTKYTLDDLVAALRHLSTS-RDVIKEILITICGTTEE---	160
α	PDCV	TEFFKFDD---IPNVNRQYLTQYTLDDLVAALRHLSTS-KDVIQEILITMCGTPED---	160
	FCoV_65F	HDFFLYKEGRCEFGNVARKDLTKYTMDLVCYAIRNFDEKNCVLEKILVTLGACNES---	151
α	HCoV_229E	HDFFTWHEGRTIYGNVSRQDLTKYTMDLVCYALRNFDKDCGVFKEILVLTGCCNTD---	149
	HCoV_NL63	HDFFTWKDGRVIYGNVSRHNLTKYTMDLVYAMRNFDQNCVDLKEVLVLTGCCDNS---	149
γ	HKU8	HDFFWVKDGRSVYGNVSRHNLTKYTMDLVHVALRNFDKNCETLKEILVLSGACDSS---	149
	PEDV	HDFFTWKDGRAIYGNVCRKDLTEYTMDLVCYALRNFDENNCVLDVLSILIKVGACEES---	149
γ	HKU22	HDFFRFE-G---MYNISRQRLTKYTMDLVCYALRHFDPNDCDVLKEILVVKGCCDWD---	149
	IBV	HDFVFNKN---IYNI SRQRLTKYTMDLVCYALRHFDPKDCVLEKILVTVGCIEDYHPK	163
β	HCoV_OC43	HDFFTFDVDEGSRVPHIVRQDLTKYTMDLVCYALRHFDRNDCMLLCDILSIYAGCEQS---	155
	HCoV_HKU1	HDFFTFDIDGSRVPHIVRRNLSKYTMDLVCYALRHFDRNDCSILCEILCEYADCKES---	155
β	MHV	HEFFTFDVEGSRVPHIVRKDLSKFTMDLVCYALRHFDRNDCSTLKEILLTYAECEES---	155
	MERS	HDFFI FVDKVKTPHIVRQRLTEYTMDLVYALRHFQDN--SEVLKAILVKYGCCDVT---	156
β	HKU4	HDFFI FVDKTKTPHIVRQSLTEYTMDLVYALRHFQDN--SEVLKAILVKYGCCDVT---	157
	HKU5	HDFFI FVDKVKTPHIVRQRLTEYTMDLVYALRHFQDN--SEVLKAILVKYGCCDAS---	157
β	HKU9	HDFFKFNI EGVMTPHIVSRERLTKYTMDLVYALRHFQDN--SEVLKAILVKYGCCDVT---	155
	SARS-CoV-2	HDFFKFRIDGDMVPHIVSRQRLTKYTMDLVYALRHFDEGNCDTLKEILVTVNCCDDD---	155
β	SARS-CoV	HDFFKFRVDDGDMVPHIVSRQRLTKYTMDLVYALRHFDEGNCDTLKEILVTVNCCDDD---	155

NiRAN Motif B_n

:** : :. * . *::*: * : . :*: :. : *

δ	HKU19	WF--VDGWYDPIENPTFYDEFHKLGLSILNCCVVMANKFADTCKTVGLVIGILTADNQDLGG	226	
	HKU11	WF--GDSWFDPIENPTFYREFHKLGSVLNRCVNLNANAFKACSELGIVIGILTPDNQDLLG	218	
	PDCV	WF--GENWFDPIENPSFYKEFHKLGDILNRCVNLNANKFASACIDAGLVIGILTPDNQDLLG	210	
	FCoV_65F	-FFENKDWFDPVENEAIHEVYARLGP IVANAMLKCVAFCDAI VEKGYIGIITLDNQDLNG	218	
	HCov_229E	-YFEMKNWFDPIENEDIHRVYAALGKVVANAMLKCVAFCDDEMVLKGVVGVLTLDNQDLNG	208	
	HCov_NL63	-YFDSKGWYDPIENEDIHRVYASLKGIVARAMLKCVALCDAMVAKGVVGVLTLDNQDLNG	208	
	HKU8	-YFDNKNWYDPIENEDIHRVYAKLGCVVANAMLKCVALCDAMVAKGVVGVLTLDNQDLNG	208	
	PEDV	-YFNKNWYDPIENEDIHRVYALLGTIVSRAMLKCVKFC DAMVEQGLVGVLTLDNQDLNG	208	
	γ	HKU22	-YFDQPNWYDPIENPDWFSLSISRLGPIQRALIKVAEFC DLMVEKGYIGVITLDNQDLNG	208
		IBV	WFEENKDWYDPIENPKYAMLAKMGPIVRRALLNAIEFGNLMVEKGYGVITLDNQDLNG	223
HCov_OC43		-YFTKKDWYDPIENPDIIINVYKKGPIFNRRALVSATEFADKLVEVGLVGVLTLDNQDLNG	214	
α	HCov_HKU1	-YFSKKDWYDPIENPDIIINIYKKGPIFNRRALLNTVIFADTLVEVGLVGVLTLDNQDLNG	214	
	MHV	-YFQKKDWYDPIENPDIIINVYKKGPIFNRRALLNTAKFADALVEAGLVGVLTLDNQDLNG	214	
	MERS	-YFENKLDWYDPIENPDIIINVYKKGPIFNRRALLNTAKFADALVEAGLVGVLTLDNQDLNG	215	
	β	HKU4	-YFDNKLWYDPIENPDIIINVYKKGPIFNRRALLNTAKFADALVEAGLVGVLTLDNQDLNG	216
		HKU5	-YFDNKLWYDPIENPDIIINVYKKGPIFNRRALLNTAKFADALVEAGLVGVLTLDNQDLNG	216
		HKU9	-YFDRKDWYDPIENPDIIIRVYHKLGETVTRKAVLSAVKMADAMVEQGLIGVLTLDNQDLNG	214
		SARS-CoV-2	-YFNKKDWYDPIENPDIIIRVYANLGERVQRALLKTVQFC DAMRNAGIVGVLTLDNQDLNG	214
	SARS-CoV	-YFNKKDWYDPIENPDIIIRVYANLGERVQRALLKTVQFC DAMRDAGIVGVLTLDNQDLNG	214	

NiRAN Motif C_N

δ	HKU19	QIYDFGDFVVTQPGNGCIEMDAYLSYIMPSMSMTHMLKCECLDD---NGSYKDYSIYQY	282	
	HKU11	QIYDFGDFIITQPGNGCVDLSSYYSYLMPIMSMTHMLKCECYDN---DGNEDYDGFQY	274	
	PDCV	QIYDFGDFIITQPGNGCVDLSSYYSYLMPIMSMTHMLKCECMD---DGNPLEYDGFQY	274	
	FCoV_65F	NFYDFGDFVKTAPGFGACVTSYYSYMPLMGMTSCLESENFVKSDFGSDYKQYDLLAY	270	
	HCov_229E	NFYDFGDFVLCPPGMGPIPYCTSYYYAMFVGMGTNCLASECFMKSDFGQDFKTFDLLKY	268	
	HCov_NL63	NFYDFGDFVLSLPMGVPCCTSYYSYMPIMGLTNCLASECFVKSDFGSDFKTFDLLKY	268	
	HKU8	NFYDFGDFITIGIPGVPLATSYYSYLMPVGMGTNCLARECFVKSEIFGSDFKTYDLELY	268	
	PEDV	DFYDFGDFTCSLKMGPIPICTSYYSYMPVGMGTNCLASECFVKSDFGDFKSYDLELY	268	
	γ	HKU22	NFYDFGDFKVLPGCGVPVTTSSYYMMPCLTACDALASERFFEFKA-TSGYKQYDLTKY	267
		IBV	KFYDFGDFQKTAPGAGVVPVFTYYSYMPPIIAMTDALAPERYFEYDV-HKGYKSYDLELY	282
HCov_OC43		KWYDFGDYVIAAPGGGVAIADSYYSYMPIMPLTMCHALDCELYV----NNAYRLFDLVQY	269	
α	HCov_HKU1	QWYDFGDFIQTAPGFGVAVADSYYSYMPMPLTMCHVLDCELFV----NDSYRQFDLVQY	269	
	MHV	QWYDFGDFVKTVPGCGVAVADSYYSYMPMPLTMCHALDSELV----NGTYREFDLVQY	269	
	MERS	KWYDFGDFVIITQPGGVAIVDSYYSYLMPVLSMTDCLAAETHRDCDF-NKPLIEWPLTEY	274	
	β	HKU4	KWYDFGDFVIITQPGGVAIVDSYYSYLMPVLSMTDCLAAETHRDCDF-NKPLIEWPLLEY	275
		HKU5	KWYDFGDFVIITQPGGVAIVDSYYSYLMPVLSMTDCLAAETHRDCDL-TKPLIEWPLLEY	275
		HKU9	QWYDFGDFIEGPAGAGVAVMDTYYSLAMPITMTNIIAAECHVSGDL-CNLKRVLDIFKY	273
		SARS-CoV-2	NWYDFGDFIQTPGSGVVPVDSYYSLMPLITLTRLALTAESHVDTDL-TKPYIKWDLLEY	273
	SARS-CoV	NWYDFGDFVQVAPGCGVPIVDSYYSLMPLITLTRLALAAESHMDADL-AKPLIKWDLLEY	273	

NiRAN Motif C_N

αCoV unique loop, PEDV 249-268

δ	HKU19	DFTDYKMELEFNKYFRHWSQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	342	
	HKU11	DFTDFKLSLFSKYFTYWRPYHPNTVDCPDDRCVHLHCANFNVLFAMCIPSTAFGNLCSQA	334	
	PDCV	DFTDFKLGLFEKYFKYWRPYHPNTVECPDDRCVHLHCANFNVLFAMCIPNTAFGNLCSRA	334	
	FCoV_65F	DFTDHKEKLFKYFKHWDRTYHPNCSDCSTSDDCIHCANFNILFAMCIPNTAFGPLVRKV	330	
	HCov_229E	DFTEHKEVLFNKYFKYWGQDYHPDCVDCHEMCIHLCSNFNTLFATTIPNTAFGPLCRKV	328	
	HCov_NL63	DFTEHKENLFNKYFKHWSFDYHPNCSDCYDDMCVILHCANFNILFATTIPGTAFGPLCRKV	328	
	HKU8	DFTEHKLGLFNKYFKHWDLDYHPNCSDCYDEMCVILHCANFNALFATTIPDTSFGPLCRKV	328	
	PEDV	DFTEHKTALFNKYFKYWGQDYHPNCVDCSDEQCIVHCANFNILFSTTIPITAFGPLCRKC	328	
	γ	HKU22	DFTEEKQLFMKYFKYWRDRTYHPNCVVECIDDRCLILHCANFNILFATLPQTAFGCLCRKV	327
		IBV	DYTEEKQDLFQKYFKYWDQYHPNCRDCSDDRCILHCANFNILFSTLVPQTSFGNLCRKV	342
HCov_OC43		DFTDYKLELFNKYFKHWSMPYHPNTVDCQDDRCIHCANFNILFSMVLNPTCFGPLVRQI	329	
α	HCov_HKU1	DFTDYKLELFNKYFKYWGMYHPNTVDCDNDRCIHCANFNILFSMVLNPTCFGPLVRQI	329	
	MHV	DFTDYKLELFNKYFKHWSMTPYHPNTCECEDDRCIHCANFNILFSMVLNPTCFGPLVRQI	329	
	MERS	DFTDYKQVLFKYFKYWDQTYHANCVNCTDDRCVHLHCANFNVLFAMTMPKTCFGPIVRKI	334	
	β	HKU4	DYTDYKIGLFNKYFKYWDQTYHPNVCNCTDDRCILHCANFNVLFSMVLNPTSGFGPIVRKI	335
		HKU5	DYTDYKIGLFNKYFKYWDQTYHPNVCNCTDDRCVHLHCANFNVLFSMVLNPTSGFGPIVRKI	335
		HKU9	YYTQFKYSLSFNKYFKYWDQTYHPNVCACADDRCILHCANFNILFSMVLNPTSGFGPLVQKI	333
		SARS-CoV-2	DFTEERLKLFDYFKYWDQTYHPNVCNCLDDRCILHCANFNVLFSVFPPTSFGPLVRKI	333
	SARS-CoV	DFTEERLCLFDYFKYWDQTYHPNVCINCLDDRCILHCANFNVLFSVFPPTSFGPLVRKI	333	

δ	HKU19	NYVD-LTYAEQNAMFEYTKRNVLPVLTQMNLK ^Y IAISAKDRAR ^T TVAGVSIISTMTNRQYHQ	581	Polymerase Motif F
	HKU11	NYVD-MTYAEQNQLFEYTKRNVLPVLTQMNLK ^Y IAISAKDRAR ^T TVAGVSIISTMTNRQYHQ	572	
α	PDCV	NYVD-MTHAEQNQLFEYTKRNVLPVLTQMNLK ^Y IAISAKDRAR ^T TVAGVSIISTMTNRQYHQ	572	
	FCoV_65F	LYYETLSYEEQDAIFALTKRNVLPMTQMNLK ^Y IAISGKARAR ^T TVGGVSLSTMTTRQYHQ	570	
	HCoV_229E	LYYESISYEEQDAIFLLTKRNVLPMTQNLN ^K YIASGKERAR ^T TVGGVSLATMTTRQFHQ	568	
	HCoV_NL63	LYYESISYEEQDALFALTKRNVLPMTQNLN ^K YIASGKERAR ^T TVGGVSLSTMTTRQYHQ	568	
	HKU8	LYYESLSYEEQDALYALTKRNVLPMTQNLN ^K YIASGKERAR ^T TVGGVSLSTMTTRQFHQ	568	
	PEDV	LYYESLSYEEQDELAYATKRNILPMTQNLN ^K YIASGKERAR ^T TVGGVSLSTMTTRQYHQ	568	
γ	HKU22	LYYESLSYAEQDQLFELTKRNVLPMTQIN ^M KYIASAKSRAR ^T TVAGVSIISTMTNRQFHQ	567	
	IBV	LYYE-MSLEEQDQLFESTKKNVLPMTIITQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQFHQ	581	
	HCoV_OC43	LYYEALSFEEQDEIYAYTKRNVLPVLTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTGRMFHQ	569	
β	HCoV_HKU1	LYYEALSFEEQNEIYAYTKRNVLPVLTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTGRMFHQ	569	
	MHV	LYYEALSFEEQDEIYAYTKRNVLPVLTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTGRMFHQ	569	
	MERS	VYYESMSYQEQDELFAVTKRNVLPMTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQYHQ	574	
	HKU4	VYYESMSYQEQDELFAVTKRNVLPMTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQYHQ	575	
	HKU5	VYYESLSYQEQDELFAVTKRNVLPMTQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQYHQ	575	
	HKU9	LYYESLSYADQDELFAVTKRNVLPMTIITQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQFHQ	573	
	SARS-CoV-2	LYYDSMSYEDQDALFAYTKRNVIPMTIITQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQFHQ	573	
SARS-CoV	LYYDSMSYEDQDALFAYTKRNVIPMTIITQMNLK ^Y IAISAKNRAR ^T TVAGVSIISTMTNRQFHQ	573		

δ	HKU19	KLLKSI SVARNQTI VIGTKFYGGWDNMLRNLMANIN ^N PKLAGW ^{DY} PKCDRSPNLLRIT	641	Polymerase Motif A
	HKU11	KMLKSI SLARNQTI VIGTKFYGGWDNMLRRLMNGINN ^P ILVGDW ^{DY} PKCDRSPNMLRIA	632	
α	PDCV	KMLKSI SLARNQTI VIGTKFYGGWDNMLRRLMNCIN ^N PI LVGDW ^{DY} PKCDRSPNMLRIA	632	
	FCoV_65F	KHLKSI AATR ^N ATVVI GTTKFYGGWDNMLKMLMRD ^V DNGLMGW ^{DY} PKCDRALPNMIRMA	630	
	HCoV_229E	KCLKSI VATR ^N ATVVI GTTKFYGGWDNMLKMLMAD ^V DDPKLMGW ^{DY} PKCDRAMP ^S MIRML	628	
	HCoV_NL63	KHLKSI VNTR ^N ATVVI GTTKFYGGWNNMLR ^T LIDGVEN ^P MLMGW ^{DY} PKCDRALPNMIRMI	628	
	HKU8	KHLKSI VNTR ^N ATVVI GTTKFYGGWDNMLRNLMD ^V DNACLMGW ^{DY} PKCDRALPNMIRMI	628	
	PEDV	KHLKSI VNTRGASVVI GTTKFYGGWDNMLKNLID ^G VEN ^P CLMGW ^{DY} PKCDRALPNMIRMI	628	
γ	HKU22	KCLKSI VNTR ^N ATVVI GTTKFYGGWDNMLRNLMR ^G VED ^P VLMGW ^{DY} PKCDRAMP ^S LLRLS	627	
	IBV	KILKSI VNTR ^N A ^P VVI GTTKFYGGWDNMLRNLIQ ^G VED ^P ILMGW ^{DY} PKCDRAMP ^N LLRIA	641	
	HCoV_OC43	KCLKSI AATR ^G V ^P VVI GTTKFYGGWDDMLR ^L L IKD ^V DN ^P VLMGW ^{DY} PKCDRAMP ^N LLRIV	629	
β	HCoV_HKU1	KCLKSI AATR ^G V ^P VVI GTTKFYGGWDDMLR ^L L IKD ^V DN ^P VLMGW ^{DY} PKCDRAMP ^N ILRIV	629	
	MHV	KCLKSI AATR ^G V ^P VVI GTTKFYGGWDDMLR ^L L IKD ^V DS ^P VLMGW ^{DY} PKCDRAMP ^N ILRIV	629	
	MERS	KMLKSI AATR ^G AT ^C VI GTTKFYGGWDFMLK ^T LYKD ^V DN ^P PHLMGW ^{DY} PKCDRAMP ^N MCRI ^F	634	
	HKU4	KMLKSI AATR ^G AT ^C VI GTTKFYGGWDFMLK ^T LYKD ^V ES ^P HLMGW ^{DY} PKCDRAMP ^N MCRI ^L	635	
	HKU5	KMLKSI AATR ^G ST ^C VI GTTKFYGGWDFMLK ^T LYKD ^V DN ^P PHLMGW ^{DY} PKCDRAMP ^N MCRI ^F	635	
	HKU9	KMLKSI AAARGASVVI GTTKFYGGWNRMLR ^T LCE ^G VDN ^P PHLMGW ^{DY} PKCDRAMP ^N LLRI ^F	633	
	SARS-CoV-2	KLLKSI AATR ^G AT ^V VIGTSK ^F YGGWHNMLK ^T VYSD ^V EN ^P PHLMGW ^{DY} PKCDRAMP ^N MLRIM	633	
SARS-CoV	KLLKSI AATR ^G AT ^V VIGTSK ^F YGGWHNMLK ^T VYSD ^V ET ^P PHLMGW ^{DY} PKCDRAMP ^N MLRIM	633		

δ	HKU19	SSLLLARKH-ACCTHSQRFYRLANCAQVLSEIVVSGNVMYV ^K PGGTS ^S SGDATTAYANSV	700	Polymerase Motif B
	HKU11	ASCLLARKH-TCCNQSQRFYRLANECQVLSEVVVSGNNLYV ^K PGGTS ^S SGDATTAYANSV	691	
α	PDCV	ASCLLARKH-TCCNQSQRFYRLANECQVLSEVVVSGNNLYV ^K PGGTS ^S SGDATTAYANSV	691	
	FCoV_65F	SAMVLGSKHIGCCTHSDRFYRLSNELAQVLTEVVHCTGGFYI ^K PGGTS ^S SGDATTAYANSA	690	
	HCoV_229E	SAMILGSKHVTCCTASDKFYRLSNELAQVLTEVVYSN ^G GGFYFKP ^G GGTTS ^S SGDATTAYANSV	688	
	HCoV_NL63	SAMVLGSKHVNCTATDRFYRLGNELAQVLTEVVYSN ^G GGFYFKP ^G GGTTS ^S SGDASTAYANSI	688	
	HKU8	SAMILGSKHVNCTNSDRYRLCNELAQVLTEVVYSN ^G GGFYMKP ^G GGTTS ^S SGDATTAYANSV	688	
	PEDV	SAMILGSKHTCCSSTDRFFRLCNELAQVLTEVVYSN ^G GGFYLPK ^G GGTTS ^S SGDATTAYANSV	688	
γ	HKU22	ASLILARRHKGCCD ^W NERIYRLANEAQVLSEVALSNGGLYV ^K PGGTS ^S SGDATTAYANSA	687	
	IBV	ASLVLARKHTNCTWSERVYRLYNECAQVLSETVLATGGIYV ^K PGGTS ^S SGDATTAYANSV	701	
	HCoV_OC43	SSLVLARKHETCCSQSDRFYRLANCAQVLSEIVMCGGCYV ^K PGGTS ^S SGDATTAFANSV	689	
β	HCoV_HKU1	SSLVLARKHEFCCSHGDRFYRLANCAQVLSEIVMCGGCYV ^K PGGTS ^S SGDATTAFANSV	689	
	MHV	SSLVLARKHDS ^C CSHTDRFYRLANCAQVLSEIVMCGGCYV ^K PGGTS ^S SGDATTAFANSV	689	
	MERS	ASLILARKHGTCTTRDRFYRLANCAQVLSEYVLCGGGYV ^K PGGTS ^S SGDATTAYANSV	694	
	HKU4	ASLILARKHSTCCTNSDRFYRLANCAQVLSEYVLCGGGYV ^K PGGTS ^S SGDATTAYANSV	695	
	HKU5	ASLILARKHSTCCTNTDRFYRLANCAQVLSEYVLCGGGYV ^K PGGTS ^S SGDATTAYANSV	695	
	HKU9	ASLILARKHSTCCNASERFYRLANCAQVLSEMVLCGGGYV ^K PGGTS ^S SGDSTTAYANSV	693	
	SARS-CoV-2	ASLVLARKHHTCCSLSHRFYRLANCAQVLSEMVLCGGSLYV ^K PGGTS ^S SGDATTAYANSV	693	
SARS-CoV	ASLVLARKHNTCCNLSHRFYRLANCAQVLSEMVLCGGSLYV ^K PGGTS ^S SGDATTAYANSV	693		

δ	HKU19	-NKKGKVFYVLLDYIRKLANELQEGIMDAFQTSTDTSYINNFTVENFYSDMYAKAPVLQ	938	
	HKU11	-PIKGGVFYLLLDYIRILAQELQDGI LDTFQSM TMSYVNNFVQEA FYAQMYEQSPTLQ	929	
	PDCV	-PIKGGVFYLLLDYIRVLAQELQDGI LDFQS L TMSYVNNFMNEAFY AQMYEQSPTLQ	929	
α	FCoV_65F	KPAYQKVFYALLDWVKHLQKTLNAGILDSFSVTMLEDGQDKFWSEEFYASLYEKSTVLQ	929	
	HCoV_229E	KPEYRKVFYALLDWVKHLNKTLENGVLESFSVTLLDEQESKFWDSEFYASMYEKSTVLQ	927	
	HCoV_NL63	NSEYRKVFYVLLDWVKHLNKNLENGVLESFSVTLLDNQEDKFWCEDFYASMYENSTILQ	927	
	HKU8	NPEYRKVFYVLLDWVKHLNNTLNQGVLESFSVTLLDASSKFWDSEFYANLYEKSAVLQ	927	PEDV nsp12
γ	PEDV	NPEYKVFYVLLDWVKHLYKTLNAGVLESFSVTLLDSTAKFWDES FYANMYEKSAVLQ	927	V894
	HKU22	NEEYKVFYVLLSYIQTLYQRLSNM LMDYSFV MNI DTSSKFWEEDFYRQMYESSPTLQ	926	
β	IBV	NEEYKVFVLLSYIRKLYQELSNM LMDYSFV MDI DRGSKFWEQEFYENMYRAPPTLQ	940	
	HCoV_OC43	NEEYQKVFVRVYLAYIKKLYNDLGNQILDSYSVILSTCDGQKFTDES FYKMYLRSAVMQ	928	
	HCoV_HKU1	NEEYQKVFVRVYLEYIKKLYNDLGTQILDSYSVILSTCDGLKFTDES FYKMYLRSAVMQ	928	
	MHV	NPEYQNVFRVYLEYIKKLYNDLGNQILDSYSVILSTCDGQKFTDET FYKMYLRSAVLQ	928	
	MERS	DIEYQNVFVVYLQYIEKLYKDLTG HMLDSYSVMLCGDNSAKFWEEAFYRDLYSSPTTLQ	933	
β	HKU4	DTEYQNVFVVYLQYIEKLYKDLTG HMLDSYSVMLCGD DSAKFWEEGFYRDLYSSPTTLQ	934	
	HKU5	DPEYQNVFVVYLQYIEKLYKDLTG HMLDSYSVMLCGDNSAKFWEE SFYRDLYTAPTTLQ	934	
	HKU9	DPEYQNVFVWYLYQYIKKLHEELTG HLLDTYSVMLASDNASKYWEVDFYENMYMESATLQ	932	
	SARS-CoV-2	NQEYADV FHLYLQYIRKLHDEL TG HMLDMSVMLTNDNTSRYWEP E FYEAMYTPHTVLQ	932	
	SARS-CoV	NQEYADV FHLYLQYIRKLHDEL TG HMLDMSVMLTNDNTSRYWEP E FYEAMYTPHTVLQ	932	
		.** * :. . * . * :. :. :. :. :. :. :. :. :. :. :. :.		

Figure S6, Multiple sequence alignment of coronavirus nsp12s. Alignment of nsp12 from the alpha- (FCoV-65F, HCoV-229E, HCoV-NL63, HKU8, PEDV), beta- (HCoV-OC43, HKU1, MHV MERS, HKU4, HKU5, HKU9, SARS-CoV, SARS-CoV-2), gamma (HKU22, IBV), and delta- (HKU19, HKU11, PDCV) coronavirus genera. Global alignment was done using Clustal Omega (3). Residues marked with “*” are conserved, “.” are very similar, and “:” are moderately similar residues.

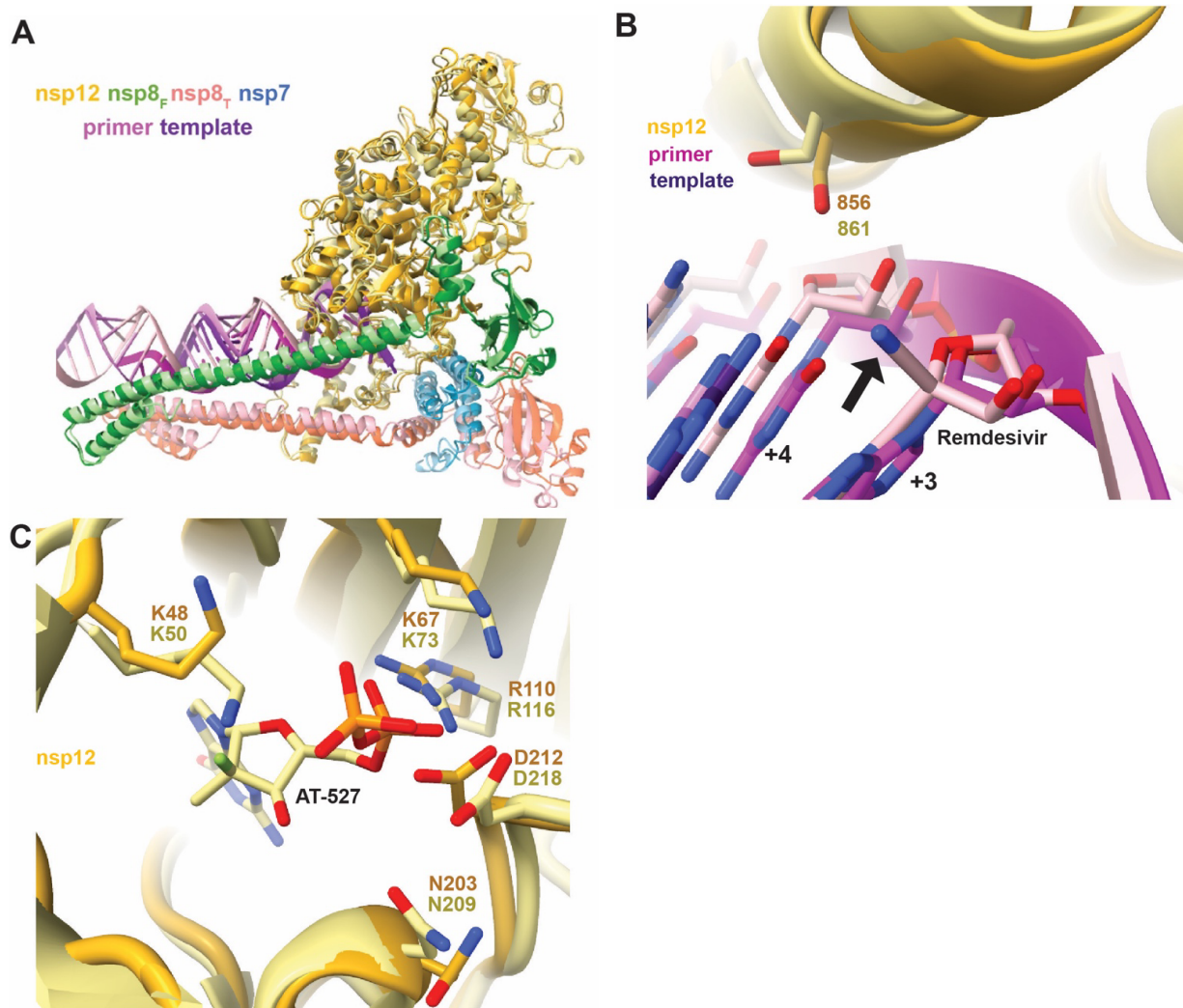


Figure S7: Possible cross-effectiveness of CoV antivirals. In each figure PEDV is shown in darker shades, superimposed are SARS-CoV-2 models in lighter, matching colors. **A)** Superimposition of complete PEDV and SARS-CoV-2 models. **B)** The antiviral Remdesivir's 1'-cyano group (black arrow) is believed to clash with SARS-CoV-2 nsp12 S861 in the +4 extension position. Remdesivir incorporated into a nascent primer at +3 is shown (PDB ID: 7B3C). **C)** The dual action antiviral AT-527 was shown to bind and inhibit the NiRAN domain of SARS-CoV-2 nsp12 (PDB ID: 7ED5). Several residues important for drug binding are conserved in PEDV nsp12, shown are K48, K67, R110, N203, and D212. Superimposition of models was achieved using the Matchmaker function in ChimeraX (4).

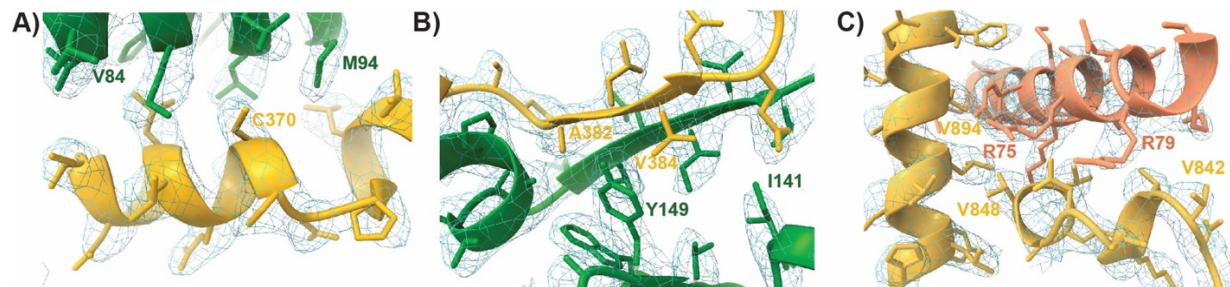


Figure S8, Model and density for mutated protein interfaces: Shown are the protein interfaces targeted for site-directed mutagenesis experiments. Interfaces tested include the nsp12 : nsp8_F N-terminal domain (**A**), nsp12 : nsp8_F head domain (**B**), and nsp12 : nsp8_T tail domain (**C**). In each image nsp12 is shown in gold, nsp8_F in blue, and nsp8_T in coral red. The map and corresponding coordinates of the complex containing nsp8_T (PDB ID: 8URB, EMD ID: 42488) were used for this figure.

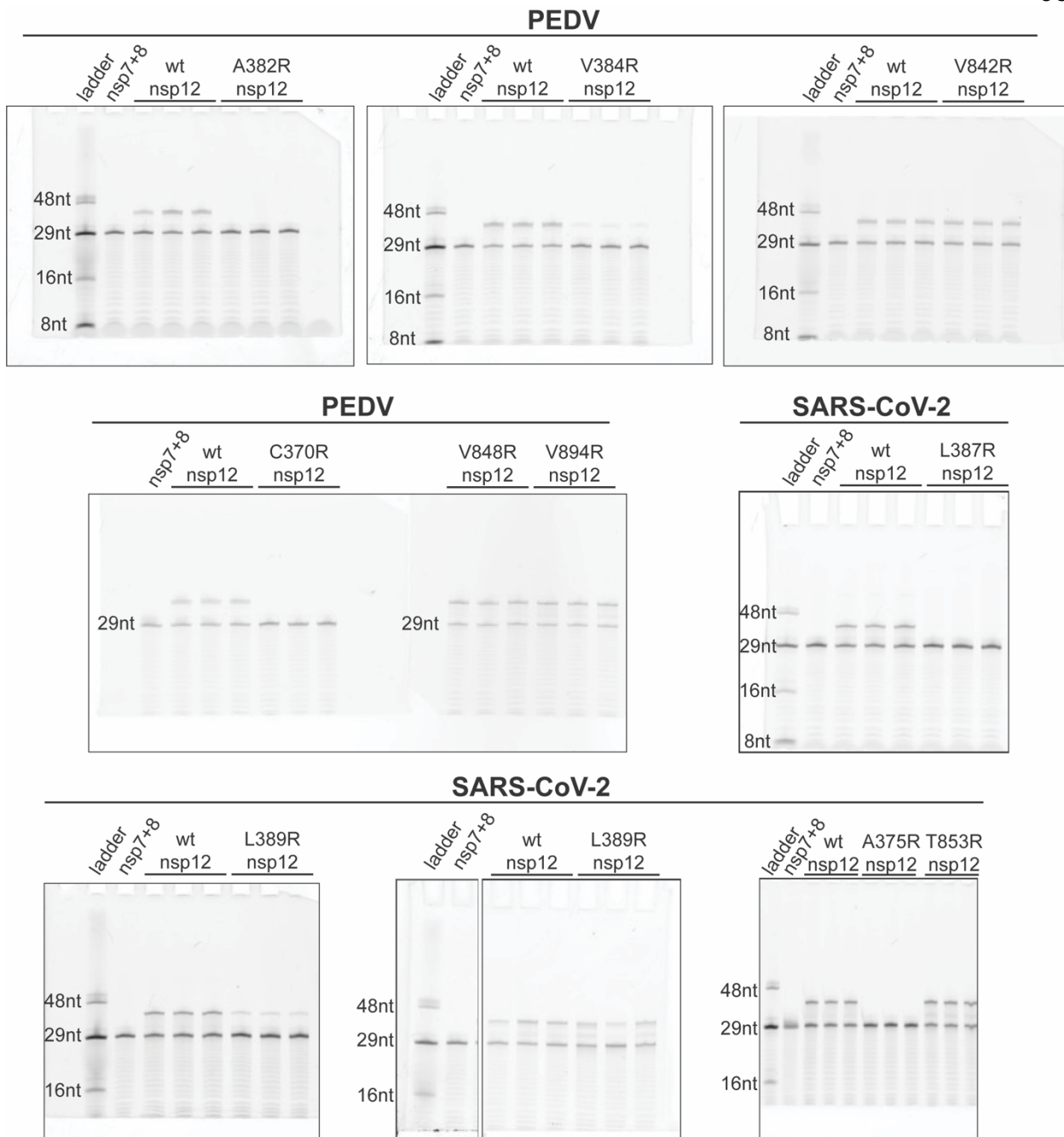


Figure S9, Full gel images for mutant nsp12 primer-extension assays: Primer extension results for complexes with mutant nsp12s for PEDV and SARS-CoV-2. For each experiment wildtype (wt) and mutant polymerase reactions were done in triplicate. Viral replication factors, nsp7 and nsp8, were always provided in excess to nsp12 (wt or mutant). Each experiment included one negative control reaction lacking nsp12.

Supplemental References:

1. Tan, Y.Z., Baldwin, P.R., Davis, J.H., Williamson, J.R., Potter, C.S., Carragher, B. and Lyumkis, D. (2017) Addressing preferred specimen orientation in single-particle cryo-EM through tilting. *Nat Methods*, **14**, 793-796.
2. Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, **14**, 290-296.
3. Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J. *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*, **7**, 539.
4. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.A.-O. and Ferrin, T.A.-O. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, **30**, 70-82.

**Chapter 3: A genus-specific nsp12 region impacts polymerase assembly in
*Alpha- and Gammacoronaviruses***

Authors: Peter J. Hoferle¹, Thomas K. Anderson¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

P.J.H. conducted the experiments to produce Figures 1A, 2, and 3.

T.K.A. performed the cryo-EM work for this project.

P.J.H. built the protein model (Fig 1B, C, D) with help from T.K.A. and R.N.K.

P.J.H. and R.N.K. wrote this chapter with input and editing by T.K.A.

Abstract:

Coronavirus relevancy has surged over the past 20 years as they have a propensity for spillover into humans from animal reservoirs resulting in pandemics such as COVID-19. The diversity within the *Coronavirinae* subfamily and high infection frequency in animal species worldwide creates a looming threat that calls for research focused across all genera within *Coronavirinae*. We sought to contribute to the limited structural knowledge within the *Gammacoronavirus* genera and determined the structure of the viral core replication-transcription complex (RTC) from Infectious Bronchitis Virus (IBV) using single-particle cryo-EM. Comparison between our solved IBV structure with published RTC structures from other *Coronavirinae* genera revealed structural differences across genera. Using *in vitro* biochemical assays, we characterized these differences and revealed their differing involvement in core RTC formation across different genera. Our findings stress the value of cross-genera *Coronavirinae* studies, as they elucidate genera specific deviations in coronavirus genome replication. A broader knowledge of coronavirus replication will better prepare us for future coronavirus spillovers.

Introduction:

Coronaviruses belong to the *Nidovirales* order of positive-sense RNA viruses. Within *Nidovirales*, this diverse subfamily of viruses is divided into four genera: the *Alpha-*, *Beta-*, *Gamma-* and *Deltacoronaviruses* (1). In 1931 the *Gammacoronavirus* infectious bronchitis virus (IBV) was the first coronavirus to ever be discovered (2). Subsequently, several additional members of the subfamily have been characterized including the human seasonal *Betacoronaviruses* HKU1 and OC43 and *Alphacoronaviruses* NL63 and 229E (1). Since 2002, three animal *Betacoronaviruses* have crossed into humans and caused disease outbreaks: SARS-CoV in 2002, MERS-CoV in 2012 and SARS-CoV-2 in 2019 (3-5). The emergence of SARS-CoV-2, the causative agent of COVID-19, led to a global pandemic that has resulted in large losses of life and significant burdens on both healthcare and the economy. In 2018, a recombinant canine-feline *Alphacoronavirus*, CCoV-HuPn-2018, was isolated from human patients hospitalized with pneumonia (6). Although CCoV-HuPn-2018 is currently incapable of efficiently infecting humans, it is poised as a preemergent human pathogen (7,8). The *Gamma-* and *Deltacoronavirus* genera contain numerous avian coronaviruses, and while no avian-to-human spillovers from these genera have been reported they, like other avian viruses, pose a consistent threat to humans due to the close contact of human and avian populations. This threat highlights the need for better characterization and monitoring of these animal coronaviruses.

The *Gammacoronavirus* genus is subdivided into three subgenera: *Igacovirus*, *Brangacovirus* and *Cegacovirus* (9). Igacoviruses are currently recognized to have three species including avian coronavirus, avian coronavirus 9203 and duck coronavirus 2714

(9). Isolates of IBV fall into both avian coronavirus and avian coronavirus 9203 species while duck coronaviruses are typically found in wild birds (10). Infection of chickens with IBV typically initiates in the respiratory tract and some strains can additionally infect the reproductive tract and kidneys. Infection of the reproductive tract can lead to a decrease in egg quality while infection of the kidneys may lead to nephritis and death (11,12). Respiratory tract infection may weaken the immune system permitting secondary bacterial pneumonia (13). Having a high prevalence in most parts of the world, IBV has been an immense economic burden on the poultry industry. Despite extensive vaccination campaigns against IBV, the large genetic diversity of the virus arising from mutation and recombination creates difficulties in providing broad protection from IBV infection (14).

Coronavirus genomes encode numerous structural and non-structural proteins used to replicate viral genomes, assemble new virions, and interact with the infected host cell (15). The 5' two-thirds of the viral genome encodes the viral non-structural proteins (nsps) responsible for viral RNA replication and transcription. These nsps are the products of polyprotein cleavage and are encoded within two open reading frames: ORF1a and ORF1b, with ORF1b accessed by -1 programmed ribosomal frameshifting at the end of ORF1a to produce either the pp1a or pp1ab polyproteins (16). Across coronavirus genera, ORF1a/b have similar organizations and cleavage products to assemble the necessary machinery for viral RNA synthesis. Within the functionally conserved suite of nsps, nsp12 encodes the RNA-dependent RNA polymerase as well as a second active site for a nucleotidyltransferase (17,18). For viruses of *Alpha-* and *Betacoronavirus*, nsp12 requires the replication factors nsp7 and nsp8 for robust RNA synthesis activity (19,20). These three nsps form the core polymerase complex that can perform processive RNA synthesis

in vitro (1). Much of the work to characterize coronavirus polymerases has focused on betacoronaviruses leaving members of other genera relatively understudied (21). One study of IBV polymerase has indicated an interaction of nsp12 with nsp8 though without a demonstration of polymerase activity or molecular descriptors (22).

Structural studies of coronavirus polymerase complexes have largely focused on complexes from SARS-CoV-2 with limited polymerase structures from SARS-CoV and the *Alphacoronavirus* porcine epidemic diarrhea virus (PEDV) (20,21,23). These structures have revealed similar nsp12 active site architectures and requirements for nsp7 and nsp8 replication factors. Expanding beyond this dataset dominated by structures of *Betacoronavirus* polymerases affords the opportunity to examine unique features of coronavirus polymerases across the subfamily while also identifying conserved mechanisms between these divergent viruses. Here, we use cryo-electron microscopy to solve the structure of the IBV polymerase complex, the first such structure from the *Gammacoronavirus* genus. We identified a genus-specific nsp12 loop that in PEDV and IBV contacts a subunit of nsp8. Subsequent biochemical analyses demonstrate the importance of this interaction and point to the potential of genus-specific polymerase complex assembly pathways. Continued investigation of coronavirus polymerases across this diverse group of viruses will aid in the development of broad-spectrum antiviral therapeutics and inform conserved pathways for viral polymerase function.

Materials and Methods:

DNA Constructs

All IBV, PEDV, and SARS-CoV-2 nsp gene sequences were codon optimized (Genscript). Sequences for IBV proteins originate from GenBank sequence

QWC71293.1. SARS-CoV-2 protein sequences originate from GenBank sequence UHD90671.1. PEDV protein sequences originate from GenBank sequence AKJ21892.1. IBV nsp7 and PEDV nsp7 were cloned into the pET46 vector with C-terminal TEV protease cleavage site and hexahistidine tag. IBV nsp8 was cloned into pET45 vector with an N-terminal hexahistidine tag and TEV protease cleavage site. IBV nsp12, PEDV nsp8 and SARS-CoV-2 nsp7 and nsp8 were cloned into pET46 vectors with N-terminal hexahistidine tags and TEV protease cleavage sites. SARS-CoV-2 and PEDV nsp12 were cloned into pFastBac vectors with C-terminal TEV cleavage site and Strep II tags. Mutant nsp12 vectors were made by performing site-directed mutagenesis on the wildtype nsp12 vectors. The sequences of all open reading frames in plasmids were confirmed using Sanger sequencing.

Protein Expression

Nsp7 and nsp8 were expressed in Rosetta 2pLysS *Escherichia coli* (*E. coli*) cells (Novagen). Cultures were grown at 37°C until they reached an OD₆₀₀ of 0.6-0.8 where they were induced with IPTG (isopropyl β-D-1-thiogalactopyranoside) at a final concentration of 500 μM and incubated overnight at 16°C. Bacterial cells were pelleted and resuspended in wash buffer (10 mM Tris-Cl, 300 mM sodium chloride, 30 mM imidazole, 2 mM dithiothreitol (DTT), pH 8). Cells were then lysed using a microfluidizer (Microfluidics) and lysate was cleared using centrifugation and filtration. Lysate supernatant was used to batch bind to Ni-NTA beads (Qiagen) for 30 minutes before loading onto a gravity column. Beads were washed with wash buffer, then protein was eluted from beads using elution buffer (10 mM Tris-Cl, 300 mM sodium chloride, 300 mM imidazole, 2 mM DTT, pH 8). Eluted proteins were buffer exchanged by dialysis (10 mM

Tris-Cl, 300 mM sodium chloride, 2 mM DTT, pH 8) while cleaving off the tag with Tobacco Etch viral (TEV) protease (1% w/w) at 4°C overnight. Proteins were passed back over a Ni-NTA column, collecting the flowthrough containing the cleaved protein sample. Protein was concentrated, then loaded onto Superdex 200 10/300 Increase GL (Cytiva) for size exclusion (25 mM Tris-Cl, 300 mM sodium chloride, 2 mM DTT, pH 8). Protein peak fractions were pooled and concentrated, then aliquoted and flash-frozen with liquid nitrogen. Proteins were stored at -80°C until use. See also **Fig. S1**.

IBV nsp12 was expressed and purified using the same protocol as above but with alternate buffers. Ni-NTA wash buffer contained 25 mM sodium-HEPES, 300 mM sodium chloride, 30 mM imidazole, 1 mM magnesium chloride, 2 mM DTT, pH 7.5. Ni-NTA elution buffer contained 10 mM sodium-HEPES, 300 mM sodium chloride, 300 mM imidazole, 1 mM magnesium chloride, 2 mM DTT, pH 7.5. Dialysis buffer contained 10 mM sodium-HEPES, 300 mM sodium chloride, 1 mM magnesium chloride, 2 mM DTT, pH 7.5. Size exclusion buffer contained 25 mM sodium-HEPES, 300 mM sodium chloride, 100 µM magnesium chloride, 2 mM tris(2-carboxyethyl)phosphine (TCEP), pH 7.5. See also **Fig. S1**.

PEDV and SARS-CoV-2 nsp12 pFastBac vectors were transformed into DH10Bac *E. coli* to generate recombinant Bacmid plasmids. Bacmid plasmids were transfected into Sf9 cells to produce baculovirus stocks that were then amplified twice before being used to infect Sf21 cells. After two days of incubation at 27°C, infected Sf21 cells were pelleted, resuspended in wash buffer (25 mM sodium-HEPES, 300 mM sodium chloride, 1 mM magnesium chloride, 2 mM DTT, pH 7.4) with an added 143 µL of biolock and lysed with a microfluidizer. Lysed cells were cleared via centrifugation and filtration. Lysates were

bound to streptactin beads (IBA) in batch for 30 minutes then loaded onto a gravity column. Beads were washed with wash buffer and proteins eluted with elution buffer (strep wash buffer with additional 2.5 mM desthiobiotin). Proteins were then concentrated, and further purified via SEC on a Superdex 200 10/300 Increase GL column with SEC buffer (25 mM sodium-HEPES, 300 mM sodium chloride, 100 μ M magnesium chloride, 2 mM TCEP, pH 7.4). Protein peak fractions were pooled, concentrated, aliquoted and then flash-frozen with liquid nitrogen. Aliquoted samples were stored at -80°C until use. See also **Fig. S1**.

RNA Substrate Preparation

RNA primers with 5' fluorescein tags (6-FAM) were annealed to longer template RNA substrates. Formation of duplex RNAs was carried out in RNA annealing buffer (2.5 mM potassium chloride, 2.5 mM HEPES, 0.5 mM magnesium chloride, pH 7.4), with a primer:template ratio of 1:1.2. After mixing, samples were heated at 95°C for 5 minutes, then slowly cooled until reaching room temperature. Annealed substrates could be used immediately or stored at -20°C .

RNA Primer for *in vitro* assays:

5' – CAUUCUCCUAAGAAGCUAUUAAAAUCACA– 3'

RNA Template for *in vitro* assays:

5' – AAAAAGGGUUGUGAUUUUAAUAGCUUCUAGGAGAAUG– 3'

RNA Primer for structure determination:

5' – CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU– 3'

RNA Template for structure determination:

5' – CAGUGUCAUGGAAAAACAGAAAAUCUGUGAUUUUAAUAGCUUCUAGGAGAAUG– 3'

Primer Extension

Primer extension assays were carried out in 20 μ L volumes with final buffer concentrations of 10 mM Tris-Cl pH 8, 2 mM magnesium chloride, 1 mM DTT and either 10 mM sodium chloride (PEDV and SARS-CoV-2) or 100 mM K-Glu (IBV). Nsp7 and nsp8, and nsp12 were combined at final concentrations of 1.5 μ M and 500 nM respectively. After combining, proteins were incubated together at 25 °C for 15 minutes followed by the addition of duplex RNA substrate (250 nM). After another 15-minute incubation at 25°C, 500 μ M of each ribonucleotide was added and varying reaction conditions were allotted for extension of primer (1 min at 25°C for SARS-CoV-2, 1 min at 30°C for PEDV, 30 min at 30°C for IBV). Reactions were then quenched by adding 2X reaction volume of denaturing RNA gel loading buffer (95% formamide (v/v), 2 mM EDTA, and 0.75 mM bromophenol blue). Quenched reactions were heated at 95°C for 15 minutes, then loaded on a denaturing urea-PAGE gel (8 M urea, 15% polyacrylamide) and run in TBE running buffer (89 mM Tris-Cl, 89 mM boric acid, 2 mM EDTA, pH 8.3). Gels were imaged using a GE Typhoon FLA 9200 scanner, using FAM tag excitation at 470 nm and measuring emission at 530 nm. The bands were analyzed using ImageJ (24).

Electrophoretic mobility shift assay

Electrophoretic mobility shift assays (EMSA) were carried out in 20 μ L reaction volumes in buffer conditions of 10 mM Tris-Cl pH 7.4, 2 mM magnesium chloride, 1 mM DTT and either 10 mM sodium chloride (PEDV and SARS-CoV-2) or 10 mM potassium glutamate (IBV). Proteins were combined at final concentrations of nsp7 (3 μ M), nsp8 (3 μ M), and nsp12 (1 μ M). Proteins were mixed and allowed to incubate at 25°C for 15 minutes. RNA substrate was added (250 nM) and the reaction incubated for an additional

15 minutes at 25°C. 10X non-denaturing gel loading buffer (10 mM Tris-Cl, 1 mM EDTA, 50% (v/v) glycerol, 0.75 mM bromophenol blue) was added to the reactions and samples were run on a 4.5% non-denaturing PAGE gel in TBE running buffer (89 mM Tris-Cl, 89 mM boric acid, 2 mM EDTA, pH 8.3). Gels were scanned using a Typhoon imager scanning for FAM fluorescence. Bands were quantitated using ImageJ (24).

Specimen Preparation for cryoEM

IBV complexes were initially assembled at a total protein concentration of 2 mg/mL in cryoEM freezing buffer (10 mM Tris-Cl pH 8, 100 mM K-Glutamate, 2 mM MgCl₂, and 1 mM DTT). Proteins and RNA were mixed at a ratio of 2:3:1:1.2 nsp7:nsp8:nsp12:RNA. To assemble the complexes, proteins were diluted in freezing buffer then combined and incubated at 25°C for 15 minutes before RNA was added and incubated for another 15 minutes at 25°C. After assembly, complexes were concentrated to 4 mg/mL total protein using ultrafiltration with a 100 kDa molecular weight cutoff. Samples were stored on ice prior to grid freezing.

Samples were frozen on UltraAuFoil R1.2/1.3 300 mesh grids (Quantifoil) using a Vitrobot Mark IV (ThermoFisher Scientific). Grids were freshly glow discharged using a GloQube Plus (Quorum) for 20 seconds with a current of 20 mA in an air atmosphere, creating a negative surface charge. Immediately prior to blotting, 0.5 µL of 3-([3-cholamidopropyl] dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO) was added to 3 µL of sample (6 mM final [CHAPSO]). 3 µL of sample+CHAPSO was spotted onto grids before double-sided blotting and vitrification in liquid ethane. Vitrobot chamber conditions were set to 100% humidity and 4°C.

CryoEM Data Collection, Processing, and Model Building

EPU (ThermoFisher Scientific) was used for data collection on a Talos Arctica 200 keV transmission electron microscope (ThermoFisher Scientific). Movies were collected using a K3 direct electron detector (Gatan) in CDS mode. A GIF quantum energy filter was used with a slit width of 20 eV. Data was collected with no stage tilt at a magnification of 79,000x with a pixel size of 1.064 Å, and a defocus range of -0.5 to -2.0 μm with a step size of 0.5 μm. Total dose per a movie was 60 e⁻/Å².

Data was processed using cryoSPARC v4.3.0 (25). After patch motion correction and CTF estimation 2,633,255 particles were picked using blob picker and extracted at a box size of 256 pixels. Particles were subjected to multiple rounds of 2D classification before three *ab initio* models were generated. Particles were classified by heterogeneous refinement using the three *ab initio* models as initial models. Output maps and classified particle stacks from heterogeneous classification were used as inputs for non-uniform refinement. Particles from the class that resembled a polymerase complex were further classified by producing two *ab initio* models with them that were then used for heterogeneous refinement. The final reconstruction was produced using non-uniform refinement with 179,183 particles (**Fig S2, S3, Table S1**).

To build a starting coordinate model, we used AlphaFold to create the IBV nsp12 RdRP and docked nsp7, nsp8, and duplex RNA from model 6YYT into our cryoEM reconstruction using ChimeraX (26-28). Model building and sequence changes were performed in Coot (29). Iterative real-space refinement in Phenix and model building and adjustments using both ISOLDE and Coot was done to generate the final coordinate model (29-31).

Results:**IBV shares replication factor requirements for RNA binding and synthesis with *Alpha-* and *Betacoronaviruses***

Recombinantly expressed and purified IBV nsp7, nsp8 and nsp12 combined with a fluorescently labeled RNA primer/template pair altered the mobility of the RNA on native-PAGE in a manner demonstrating that both nsp7 and nsp8 are required for RNA binding to nsp12 (**Fig. S4**). Adding nucleotides to this complex similarly demonstrated the IBV nsp12 requirement for both nsp7 and nsp8 for promoting robust RNA synthesis activity by primer extension (**Fig. 1A and S5**). The results of these assays indicate similar requirements for nsp7 and nsp8 for RNA binding and polymerase activity across the coronavirus subfamily (19,20,27).

Structure of the IBV polymerase complex

Single-particle cryo-electron microscopy was used to solve the structure of the IBV nsp7-nsp8-nsp12-RNA complex. Our cryo-EM reconstruction had a resolution of 3.5 Å (**Fig. 1B and C, S2, S3, Table S1**). Clearly visible in the map are densities for all components including nsp12, nsp7, two copies of nsp8, and the RNA substrate with a complex stoichiometry of 1:2:1 for nsp7:nsp8:nsp12 which is consistent with other published coronavirus polymerase structures (20,23,32) (**Fig. 1D**). Similar to *Alpha-* and *Betacoronavirus* polymerases one protomer of nsp8 binds the nsp12 fingers domain (nsp8_F) while a second protomer binds to nsp12 as a nsp7-nsp8 heterodimer adjacent to the nsp12 thumb domain (nsp8_T). The identification of nsp8_F is congruent with a previous biochemical study of IBV nsp12 demonstrating an interaction of nsp8 with nsp12 residues 1-400 which encompasses nearly the entirety of our observed nsp8_F binding site on nsp12

(22). As previously observed in coronavirus polymerase structures bound to duplex RNAs, the N-terminal extensions of each nsp8 form long helices to contact upstream double-stranded RNA extending from the polymerase active site (23,27). The IBV polymerase and nucleotidyltransferase active sites are well resolved and well conserved both in sequence and structure among SARS-CoV, SARS-CoV-2 and PEDV polymerases (SARS-CoV: 6NUR, SARS-CoV-2: 7KRP, PEDV: 8URB) suggesting the broad applicability of antiviral drugs targeting these sites (33,34).

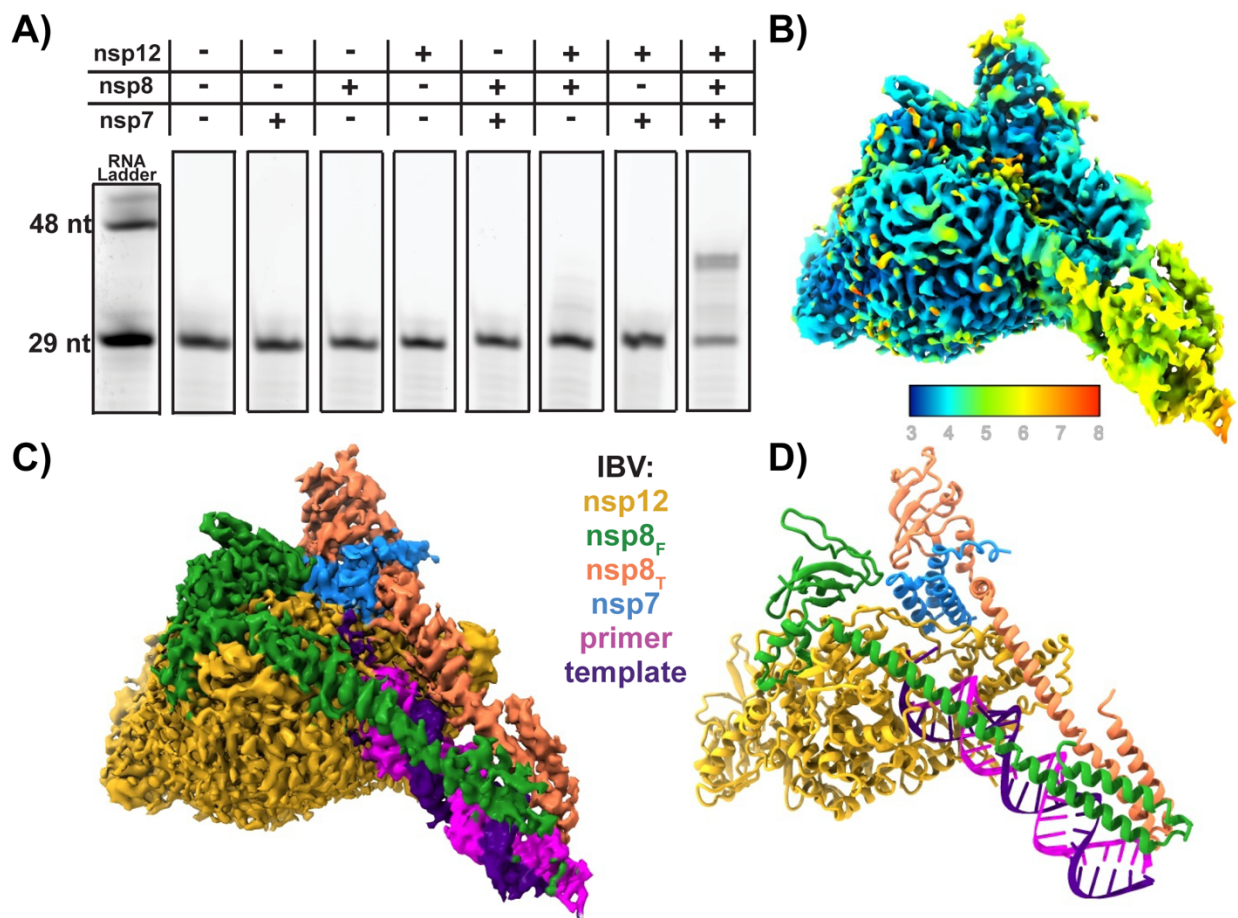


Figure 1, Structure of an active IBV polymerase complex: A) Extension of a short (29 nt) RNA primer to the length of the template RNA (38 nt) in the presence of the IBV polymerase complex. **B and C)** Cryo-EM reconstruction of the IBV polymerase complex colored by local resolution (**B**) or chain (**C**). **D)** Atomic model of the IBV polymerase complex built using the cryo-EM reconstruction.

Structurally observed insertions and deletions in the *Gammacoronavirus* polymerase complex

Despite the high sequence and structural homology of coronavirus polymerases, we identified large insertions and deletions in nsp8 and nsp12 that result in unique conformations within the IBV polymerase complex structure. Many of these regions are distal to known active sites and protein-protein interfaces and their influence on the viral polymerase remains unclear.

IBV nsp8 loop 173-181 contains an insertion not observed in other genera of coronaviruses (**Fig. S6**). This loop sequence is well conserved among *Igacovirus* nsp8s while *Brangocovirus* nsp8s contain an additional three amino acid insertion and the *Cegacovirus* nsp8s have a nine amino acid deletion. *Betacoronavirus* nsp8s possess shorter loops in this nsp8 region with *Embecovirus* members, such as Murine Hepatitis Virus, having nsp8 loops nine amino acids shorter than IBV. Similarly, *Alpha-* and *Deltacoronavirus* nsp8s have loops five and 13 amino acids shorter than IBV, respectively. This IBV nsp8 region lacks secondary structure and forms an extended loop from the nsp8 C-terminal head domains, while in SARS-CoV, SARS-CoV-2 and PEDV this loop forms a pair of short helices (**Fig. S7**). This loop is clearly visible in the IBV reconstructed density in both nsp8_F and nsp8_T though the density is weaker at the distal end of the loop particularly for nsp8_T. Additional examination of IBV nsp8_T reveals an 18° rotation in the conformation of the nsp8 head domain relative to nsp7 when compared to corresponding domains from SARS-CoV-2 (35). This rotated nsp8_T head domain is similar to the orientation of nsp8_T for PEDV (23). IBV and PEDV nsp8_T regions 122-129 (IBV) also

appear to make more extensive contacts to nsp7 α 2 while in SARS-CoV and SARS-CoV-2 equivalent nsp8_T regions appear more restricted to nsp7 α 3 (**Fig S6, S8**) (20,23,27).

In addition to nsp8, there are several insertions and deletions in IBV nsp12 particularly in the N-terminal nucleotidyltransferase domain (18). There is a large insertion in IBV nsp12 loop 67-72 when compared to *Alpha*- and *Betacoronavirus* nsp12s (**Fig. S9**). This loop sequence is well conserved in avian coronaviruses and avian coronavirus 9203 with some length polymorphisms in other *Gammacoronavirus* nsp12s. In *Alpha*- and *Betacoronavirus* nsp12, this loop is four and eight amino acids shorter, respectively, and one to two amino acids longer in *Deltacoronavirus*. This loop is positioned on the opposite side of the nucleotidyltransferase domain from the enzyme active site and is distal to known protein-binding sites for nsp7, nsp8, nsp9, and nsp13 (**Fig. S10**) (23,35,36).

There is a shortened loop in IBV nsp12 113-115 that is conserved in all *Gammacoronavirus* nsp12s except for *Cegacovirus* nsp12s which are one amino acid shorter, similar to *Deltacoronavirus* nsp12s. In contrast, *Alpha*- and *Betacoronavirus* nsp12s are four amino acids longer in this region (**Fig. S9**). This nsp12 loop lies within the nucleotidyltransferase domain but again is distant from the enzyme active site and known protein-binding sites (**Fig. S11**).

IBV nsp12 contains a four amino acid insertion in loop 156-169 compared to *Alpha*-, *Beta*- and *Deltacoronavirus* nsp12s (**Fig. S9**). In *Gammacoronavirus*, the nsp12 loop 156-169 (IBV) is conserved across avian coronavirus species while duck coronaviruses and *Brangacovirus* contain an additional 10 amino acid insertion with more divergent sequences. *Cegacovirus* nsp12s have shorter loops of similar length to the other coronavirus genera. Structurally, this region of nsp12 appears to be conserved in both

Alpha- and *Betacoronavirus* nsp12s which is unsurprising given nsp12s' moderate sequence conservation between these two genera. The insertion in IBV nsp12 loop 156-169 results in a shortened helical region and more extended loop region that extends outwards from the polymerase (**Fig. S12**). While loop sequence insertions at this position have so far only been noted for coronaviruses infecting avian species, the lack of insertions here in avian *Deltacoronavirus* and the poor representation of *Cegacovirus* sequences that infect mammals in databases warns against identifying this loop as a potential host species determinant.

Sequence and structural comparison of the IBV nsp12 loop 264-278

In the IBV nsp12 interface domain, amino acids 264-278 form a large loop that is in a dramatically different conformation when compared to the homologous region in both *Alpha-* and *Betacoronavirus* (**Fig. 2B**) (20,32). This loop extends from the interface domain to contact the head domain of nsp8_F. We define this loop as being flanked by well conserved residues L279 and L280 that form the hydrophobic pocket for nsp8_F region 103-129 and E263 that forms a well conserved salt bridge with K294 (IBV nsp12 numbering). The corresponding region in PEDV nsp12 (249-264) also extends to contact the nsp8_F head domain but adopts a different conformation to use distinct regions on both nsp12 and nsp8_F to form the protein-protein interaction (20). In contrast, the loop in *Betacoronavirus* SARS-CoV and SARS-CoV-2 nsp12s extends away from the core complex not forming any interactions with known replication factors (23,32).

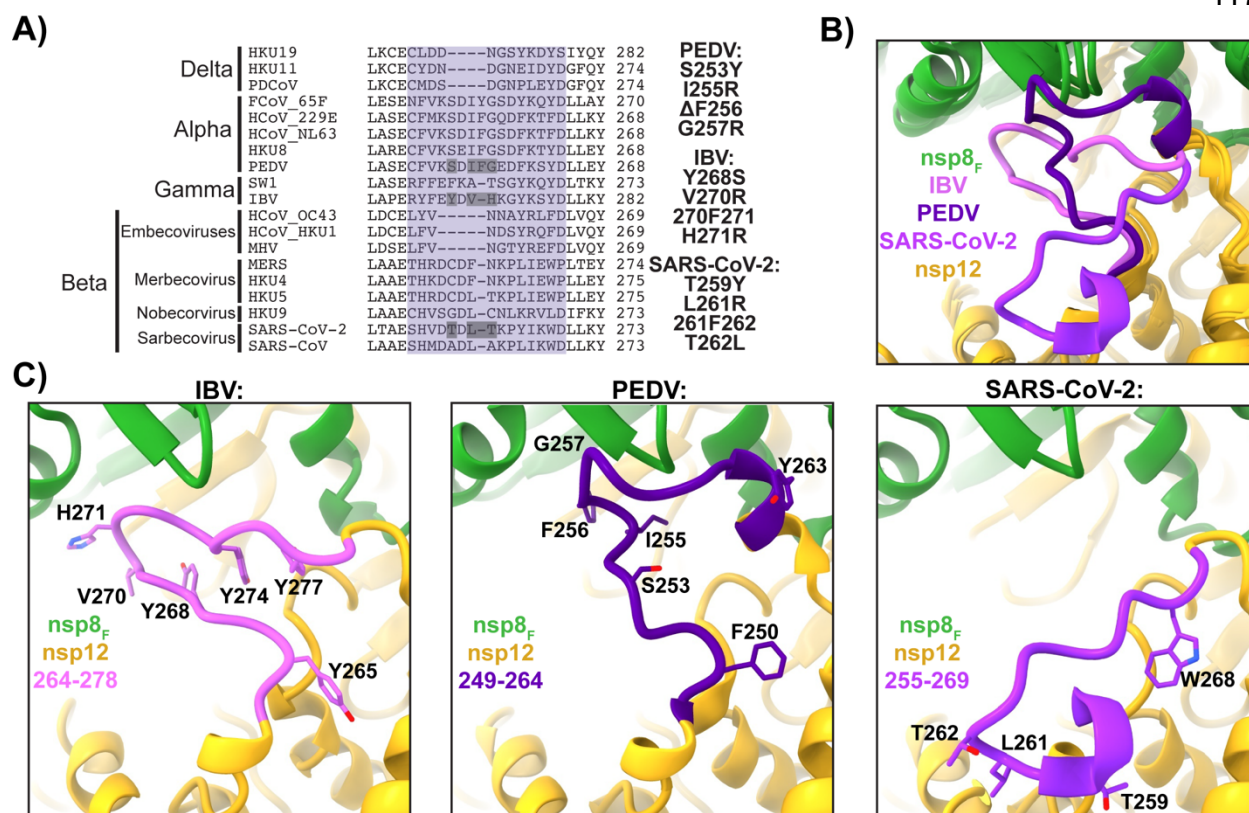


Figure 2, Altered sequence and structure of nsp12-nsp8_F interactions: A) Multiple sequence alignment of nsp12 residues 260–282 (IBV) across the coronavirus subfamily. The loop region of interest is highlighted in purple, and residues that were mutated in this study are listed to the right and have gray boxes around them. **B)** Superimposition of IBV, PEDV (8URB), and SARS-CoV-2 (7KRP) highlighting the altered conformations of the nsp12 loop (IBV residues 264–278). **C)** Individual snapshots of the nsp12 loop (IBV residues 264–278) from IBV (left), PEDV (center), and SARS-CoV-2 (right).

Examining sequence alignments for this region of nsp12, the *Betacoronavirus* subgenera *Nobecovirus*, *Sarbecovirus* and *Merbecovirus* have loop lengths similar to *Gammacoronavirus* nsp12 while *Embecovirus* nsp12s have loops that are four amino acids shorter (**Fig. 2A**). While there is no structural data to provide insight into this nsp12 region for *Deltacoronavirus* polymerases, sequence alignment indicates a shortening of this loop by three amino acids relative to IBV. Hence, we hypothesize that the nsp12 loop does not contact nsp8_F in *Deltacoronavirus* nsp12s. In contrast *Alphacoronavirus* nsp12s have a one amino insertion, adding a Phe at position 256 (PEDV numbering). PEDV nsp12 F256 lies at the apex of the nsp12 loop and packs into the hydrophobic surface between nsp12 and nsp8_F. To test the role of F256 in PEDV complex assembly we produced a recombinant PEDV nsp12 with F256 deleted (PEDV nsp12 Δ F256) which resulted in nearly a complete loss of polymerase activity using our aforementioned *in vitro* RNA primer extension assay (**Fig. 3**). Complementary to this PEDV nsp12 deletion, we inserted a Phe into the corresponding nsp12 positions of IBV (270F271) and SARS-CoV-2 (261F262). Neither of these insertions diminished the ability of the mutant polymerases to bind RNA or extend primers (**Fig. 3 and S13, S14**). In IBV nsp12, while this loop contacts nsp8_F, the 270F271 insertion would be expected to be surface exposed owing to different utilization of this nsp12 region to contact the nsp8_F head domain.

Examining the conformation of the IBV nsp12 264-278 loop, Tyr residues at 268, 274 and 277 undergo aromatic stacking and likely stabilize this unique loop conformation. To test the importance of the tyrosine stacking in maintaining the IBV nsp12 264-278 loop conformation, we created an IBV nsp12 with a Y268S mutation. IBV nsp12 Y268S showed a more than 50% reduction in polymerase activity as well as a strong defect in

RNA binding (**Fig. 3 and S13,14**). These defects in polymerase activities likely are caused by a failure to properly assemble IBV nsp8_F on nsp12 Y268S and point to the importance of this IBV nsp12 loop region for polymerase complex binding to RNA. Complementary mutations in PEDV nsp12 (S253Y) or SARS-CoV-2 nsp12 (T259Y) did not have major effects on viral polymerase activities supporting the distinct conformations and interactions of these nsp12 loop regions observed in the structural data (**Fig. 3 and Fig S13,14**). The presence of an aromatic residue at nsp12 277 (IBV numbering, SARS-CoV-2 W268, PEDV Y263) is well conserved across the coronavirus subfamily, however, these aromatic residues are placed in very different contexts within this loop region. In PEDV, nsp12-Y263 is surface exposed while in IBV nsp12-Y277 participates in the tyrosine stacking interaction that stabilizes this unique loop conformation. In contrast, SARS-CoV and SARS-CoV-2 nsp12-W268 is oriented into a hydrophobic pocket on nsp12 which may drive the diversion of this loop into the observed outwards directed conformation. Structural comparison of the SARS-CoV-2 nsp12-W268 hydrophobic pocket shows that the homologous pocket is occupied by IBV nsp12-Y265 or PEDV-F250. The functional constraint of needing to insert an aromatic residue into this hydrophobic pocket may be driving loop conformational differences and presentation of this nsp12 loop to nsp8_F in IBV and PEDV polymerase complexes.

To further examine specific interactions at the IBV nsp12 264-278 and PEDV nsp12 249-264 loop apexes with their respective nsp8_F head domains, we generated nsp12 mutants for key residues in the protein interfaces and complementary mutations in other coronavirus polymerases (**Fig. 3 and S13,14**). In addition to PEDV nsp12 F256, I255 contributes to the buried hydrophobic surface between nsp12 and the nsp8_F head

domain. A PEDV nsp12-I255R mutation reduced primer extension activity by 50% and prevented strong RNA binding to the PEDV polymerase complex. Homologous mutations to IBV nsp12 (V271R) and SARS-CoV-2 nsp12 (L261R) did not have large effects on either polymerase complex. For IBV nsp12, H271 resides at the apex of the 264-278 loop to contact the nsp8_F head domain. IBV nsp12-H271R had a 50% reduction in polymerase activity and a significant loss in RNA binding activity. Mutations to homologous positions in PEDV nsp12 (G257R) or SARS-CoV-2 nsp12 (T262L) had no effect on either polymerase primer extension or RNA binding activity. These targeted mutations highlight the distinct interactions of each genus's nsp12 loop with nsp8_F and that disrupting these interactions has genus-specific negative impacts on polymerase activity.

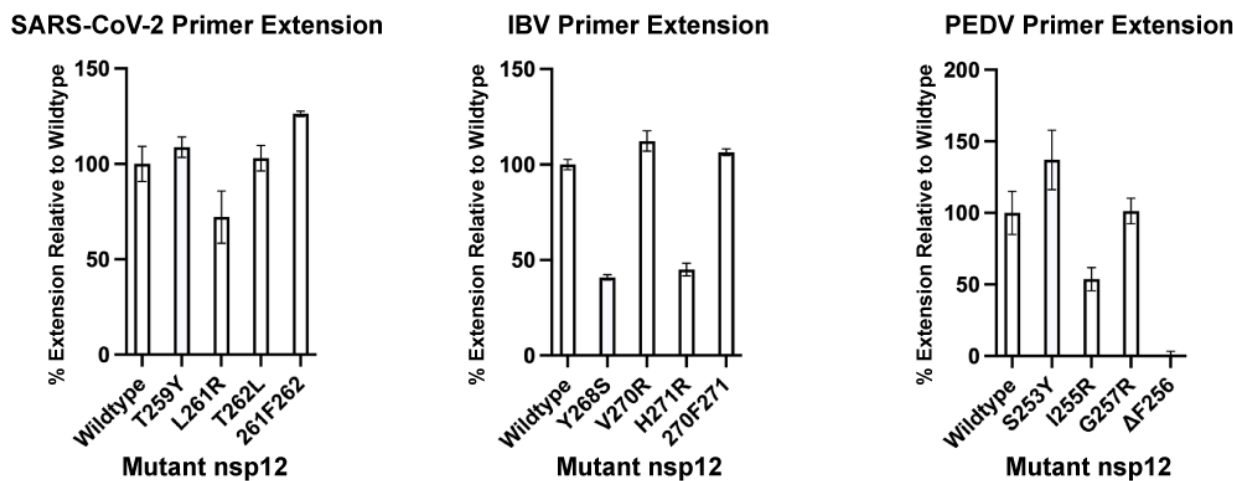


Figure 3, Mutant nsp12 primer extensions: Point mutations, insertions, and deletions of nsp12 from SARS-CoV-2 (left), IBV (center), and PEDV (right) were tested for their ability to form the core-RTC and extend RNA *in vitro*. For each coronavirus results are presented as % activity of the polymerase complex with wildtype nsp12.

Discussion:

Here we have presented the first structure of a *Gammacoronavirus* polymerase complex showing the IBV RNA polymerase bound to its essential replication factors and RNA. Structural comparisons highlight a loop in IBV nsp12 (residues 264-278) that is in an alternate conformation than previous polymerase complex structures from PEDV and SARS-CoV-2. Mutagenesis of key residues in this protein region among these three polymerase complexes were a detriment to IBV and PEDV polymerase RNA-binding and primer extension activities. The inability of these polymerases to bind RNA is likely a result of defects in the nsp12s' ability to assemble properly with replication factor nsp8_F. It has been previously shown that replication factors nsp7 and nsp8 are essential for nsp12 RNA binding and that disruptions to the nsp8_F head domain - nsp12 interaction resulted in polymerases incapable of extending primers or binding RNA despite not fully blocking nsp8_F subunit binding to the complex (19,23). We identify this nsp12 loop as a genus-specific structural feature that is functionally important for the proper assembly of *Alpha-*

and *Gammacoronavirus* polymerases. Work on SARS-CoV-2 polymerase complexes have identified the binding of nsp8_F as a rate limiting step in polymerase assembly (37). The observed altered interactions of IBV and PEDV nsp12 with nsp8_F may indicate the existence of alternate polymerase assembly pathways across diverse viruses. Our observed structural differences among viral nsps with functional consequences for polymerase activity highlight the need to consider alternate assembly and functional pathways across diverse coronavirus genera more broadly. In determined structures, the IBV nsp12 264-278 and PEDV nsp12 249-264 loops neighbor SARS-CoV-2 nsp12 P323. During the early months of the COVID-19 pandemic variant strains carrying a P323L mutation along with a D614G mutation in the viral spike rapidly rose to prominence (38). The impacts of the SARS-CoV-2 nsp12 P323L mutation on polymerase activity remain unclear but the spatial proximity of this mutation to the observed altered loop conformations in IBV and PEDV polymerases creates the possibility that this region of the nsp12 polymerase has a role in modulating the activity and assembly of the coronavirus polymerase complex across viral evolution.

References:

- 1.) Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol.* 2015;1282:1-23. doi: 10.1007/978-1-4939-2438-7_1. PMID: 25720466; PMCID: PMC4369385.
- 2.) Schalk A.F., Hawin M.C. An apparently new respiratory disease in baby chicks. *J. Am. Vet. Med. Assoc.* 1931;78:413–422
- 3.) Snijder, E.J., et al., Unique and conserved features of genome and proteome of SARS-coronavirus, an early split-off from the coronavirus group 2 lineage. *J Mol Biol*, 2003. 331: p. 991-1004

- 4.) Zaki, A.M., et al., Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, 2012. 367: p. 1814-20
- 5.) Wu, F., et al., A new coronavirus associated with human respiratory disease in China. *Nature*, 2020. 579: p. 265-269
- 6.) Vlasova AN, Diaz A, Dامتie D, Xiu L, Toh TH, Lee JS, Saif LJ, Gray GC. Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis*. 2022 Feb 11;74(3):446-454. doi: 10.1093/cid/ciab456. PMID: 34013321; PMCID: PMC8194511.
- 7.) Liu Y, Chen D, Wang Y, Li X, Qiu Y, Zheng M, Song Y, Li G, Song C, Liu T, Zhang Y, Guo JT, Lin H, Zhao X. Characterization of CCoV-HuPn-2018 spike protein-mediated viral entry. *J Virol*. 2023 Sep 28;97(9):e0060123. doi: 10.1128/jvi.00601-23. Epub 2023 Sep 7. PMID: 37676001; PMCID: PMC10537617.
- 8.) Tortorici MA, Walls AC, Joshi A, Park YJ, Eguia RT, Miranda MC, Kepl E, Dosey A, Stevens-Ayers T, Boeckh MJ, Telenti A, Lanzavecchia A, King NP, Corti D, Bloom JD, Veesler D. Structure, receptor recognition, and antigenicity of the human coronavirus CCoV-HuPn-2018 spike glycoprotein. *Cell*. 2022 Jun 23;185(13):2279-2291.e17. doi: 10.1016/j.cell.2022.05.019. Epub 2022 May 27. PMID: 35700730; PMCID: PMC9135795.
- 9.) Marchenko V, Danilenko A, Kolosova N, Bragina M, Molchanova M, Bulanovich Y, Gorodov V, Leonov S, Gudymo A, Onkhonova G, Svyatchenko S, Ryzhikov A. Diversity of gammacoronaviruses and deltacoronaviruses in wild birds and poultry in Russia. *Sci Rep*. 2022 Nov 12;12(1):19412. doi: 10.1038/s41598-022-23925-z. PMID: 36371465; PMCID: PMC9653423
- 10.)Michelle Wille, Edward C Holmes, Wild birds as reservoirs for diverse and abundant gamma- and deltacoronaviruses, *FEMS Microbiology Reviews*, Volume 44, Issue 5, September 2020, Pages 631–644, <https://doi.org/10.1093/femsre/fuaa026>
- 11.)Crinion RA, Ball RA, Hofstad MS. Abnormalities in laying chickens following exposure to infectious bronchitis virus at one day old. *Avian Dis*. 1971 Jan-Mar;15(1):42-8. PMID: 5547755.
- 12.)WINTERFIELD RW, HITCHNER SB. Etiology of an infectious nephritis-nephrosis syndrome of chickens. *Am J Vet Res*. 1962 Nov;23:1273-9. PMID: 14001258.
- 13.)Swayne D.E., Glisson J.R., Mcdougald L.R., Nolan L.K., Suarez D.L., Nair V.L. thirteenth ed. 2013. *Diseases of Poultry*. Diseases of Poultry.

- 14.) Bali K, Bálint Á, Farsang A, Marton S, Nagy B, Kaszab E, Belák S, Palya V, Bányai K. Recombination Events Shape the Genomic Evolution of Infectious Bronchitis Virus in Europe. *Viruses*. 2021 Mar 24;13(4):535. doi: 10.3390/v13040535. PMID: 33804824; PMCID: PMC8063831.
- 15.) Brian DA, Baric RS. Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*. 2005;287:1-30. doi: 10.1007/3-540-26765-4_1. PMID: 15609507; PMCID: PMC7120446.
- 16.) Brierley I, Bournsnel ME, Binns MM, Bilimoria B, Blok VC, Brown TD, Inglis SC. An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J*. 1987 Dec 1;6(12):3779-85. doi: 10.1002/j.1460-2075.1987.tb02713.x. PMID: 3428275; PMCID: PMC553849.
- 17.) Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. *Nucleic Acids Res*. 2003 Dec 15;31(24):7117-30. doi: 10.1093/nar/gkg916. PMID: 14654687; PMCID: PMC291860.
- 18.) Kathleen C. Lehmann, Anastasia Gulyaeva, Jessika C. Zevenhoven-Dobbe, George M. C. Janssen, Mark Ruben, Hermen S. Overkleeft, Peter A. van Veelen, Dmitry V. Samborskiy, Alexander A. Kravchenko, Andrey M. Leontovich, Igor A. Sidorov, Eric J. Snijder, Clara C. Posthuma, Alexander E. Gorbalenya, Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses, *Nucleic Acids Research*, Volume 43, Issue 17, 30 September 2015, Pages 8416–8434, <https://doi.org/10.1093/nar/gkv838>
- 19.) Subissi L, Posthuma CC, Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, Snijder EJ, Canard B, Imbert I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc Natl Acad Sci U S A*. 2014 Sep 16;111(37):E3900-9. doi: 10.1073/pnas.1323705111. Epub 2014 Sep 2. PMID: 25197083; PMCID: PMC4169972.
- 20.) Anderson TK, Hoferle PJ, Lee KW, Coon JJ, Kirchdoerfer RN. An alphacoronavirus polymerase structure reveals conserved co-factor functions. *bioRxiv [Preprint]*. 2023 Mar 16:2023.03.15.532841. doi: 10.1101/2023.03.15.532841. PMID: 36993498; PMCID: PMC10055115.
- 21.) Malone, B., Urakova, N., Snijder, E.J. et al. Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat Rev Mol Cell Biol* 23, 21–39 (2022). <https://doi.org/10.1038/s41580-021-00432-z>

- 22.)Tan YW, Fung TS, Shen H, Huang M, Liu DX. 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*. 2018 Jan 1;513:75-84. doi: 10.1016/j.virol.2017.10.004. Epub 2017 Oct 13. PMID: 29035788; PMCID: PMC7112110.
- 23.)Kirchdoerfer, R.N., Ward, A.B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 10, 2342 (2019). <https://doi.org/10.1038/s41467-019-10280-3>
- 24.)Schneider, C. A., Rasband, W. S., & Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature Methods*, 9(7), 671–675. [doi:10.1038/nmeth.2089](https://doi.org/10.1038/nmeth.2089)
- 25.)Punjani, A., Rubinstein, J., Fleet, D. et al. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods* 14, 290–296 (2017). <https://doi.org/10.1038/nmeth.4169>
- 26.)Jumper, J. et al. “Highly accurate protein structure prediction with AlphaFold.” *Nature*, 596, pages 583–589 (2021). DOI: [10.1038/s41586-021-03819-2](https://doi.org/10.1038/s41586-021-03819-2)
- 27.)Hillen, H.S., Kokic, G., Farnung, L. et al. Structure of replicating SARS-CoV-2 polymerase. *Nature* 584, 154–156 (2020). <https://doi.org/10.1038/s41586-020-2368-8>
- 28.)Meng EC, Goddard TD, Pettersen EF, Couch GS, Pearson ZJ, Morris JH, Ferrin TE. *Protein Sci*. 2023 Nov;32(11):e4792.
- 29.)Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*. 2010 Apr;66(Pt 4):486-501. doi: 10.1107/S0907444910007493. Epub 2010 Mar 24. PMID: 20383002; PMCID: PMC2852313.
- 30.)Croll, Tristan Ian. 2018. “ISOLDE: A Physically Realistic Environment for Model Building into Low-Resolution Electron-Density Maps.” *Acta Crystallographica Section D: Structural Biology* 74 (6):519–30. doi: [10.1107/S2059798318002425](https://doi.org/10.1107/S2059798318002425).
- 31.)D. Liebschner, P.V. Afonine, M.L. Baker, G. Bunkóczi, V.B. Chen, T.I. Croll, B. Hintze, L.W. Hung, S. Jain, A.J. McCoy, N.W. Moriarty, R.D. Oeffner, B.K. Poon, M.G. Prisant, R.J. Read, J.S. Richardson, D.C. Richardson, M.D. Sammito, O.V. Sobolev, D.H. Stockwell, T.C. Terwilliger, A.G. Urzhumtsev, L.L. Videau, C.J. Williams, and P.D. Adams. [Acta Cryst. D75, 861-877 \(2019\)](https://doi.org/10.1107/S2059798318002425).

- 32.) 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 LW, Gong P, Rao Z. Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*. 2020 Jul 23;182(2):417-428.e13. doi: 10.1016/j.cell.2020.05.034. Epub 2020 May 22. PMID: 32526208; PMCID: PMC7242921.
- 33.) Shannon, A., Fattorini, V., Sama, B. et al. A dual mechanism of action of AT-527 against SARS-CoV-2 polymerase. *Nat Commun* 13, 621 (2022). <https://doi.org/10.1038/s41467-022-28113-1>
- 34.) Gordon, Calvin J., et al. "Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency." *Journal of Biological Chemistry* 295.20 (2020): 6785-6797.
- 35.) Chen J, Malone B, Llewellyn E, Grasso M, Shelton PMM, Olinares PDB, Maruthi K, Eng ET, Vatandaslar H, Chait BT, Kapoor TM, Darst SA, Campbell EA. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*. 2020 Sep 17;182(6):1560-1573.e13. doi: 10.1016/j.cell.220.07.033. Epub 2020 Jul 28. PMID: 32783916; PMCID: PMC7386476.
- 36.) Park, G.J., Osinski, A., Hernandez, G. et al. The mechanism of RNA capping by SARS-CoV-2. *Nature* 609, 793–800 (2022). <https://doi.org/10.1038/s41586-022-05185-z>
- 37.) Campagnola G, Govindarajan V, Pelletier A, Canard B, Peersen OB. 2022. The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol* 96:e00671-22. <https://doi.org/10.1128/jvi.00671-22>
- 38.) Pachetti, M., Marini, B., Benedetti, F. et al. Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *J Transl Med* 18, 179 (2020). <https://doi.org/10.1186/s12967-020-02344-6>

Supplemental Figures:

EMDB	<i>TBD</i>
PDB	<i>TBD</i>
Microscope	Talos Arctica
Voltage (kV)	200
Detector	K3 direct electron detector (Gatan)
Dose Rate (e ⁻ /pixel/sec)	14.2
Exposure Time (sec)	4.78
Electron Exposure (e ⁻ /Å ²)	60
Frames (no.)	60
Defocus Values	-0.5, -1.0, -1.5, -2.0
Data Collection Mode	EFTEM, Counting, CDS
Nominal Magnification	79,000
Pixel Size (Å)	1.064
Symmetry Imposed	C1
Movies Collected (no.)	5,777
Initial Particle Images (no.)	2,633,225
Final Particle Images (no.)	179,183
Map Resolution (Å) – GSFSC	3.5
Initial Models Used (PDB ID)	6YYT
Non-hydrogen Atoms	12,465
Protein Residues	1,455
Nucleic Acid Residues	63
Other Atoms	2 Zn ²⁺
R.M.S. Deviations	
Bond Lengths (Å)	0.003
Bond angles (°)	0.463
MolProbity Score	1.60
Clashscore	6.61
Ramachandran Plot	
Favored (%)	96
Allowed (%)	4
Disallowed (%)	0

Table S1, Cryo-EM data collection and refinement: Information provided is for the cryoEM data collection, and processing that produced the electron density map for the IBV polymerase complex. PDB and EMDB codes are not provided as final model adjustments for PDB submission are still being done, in addition, because of this model validation statistics may be slightly altered compared after publication.

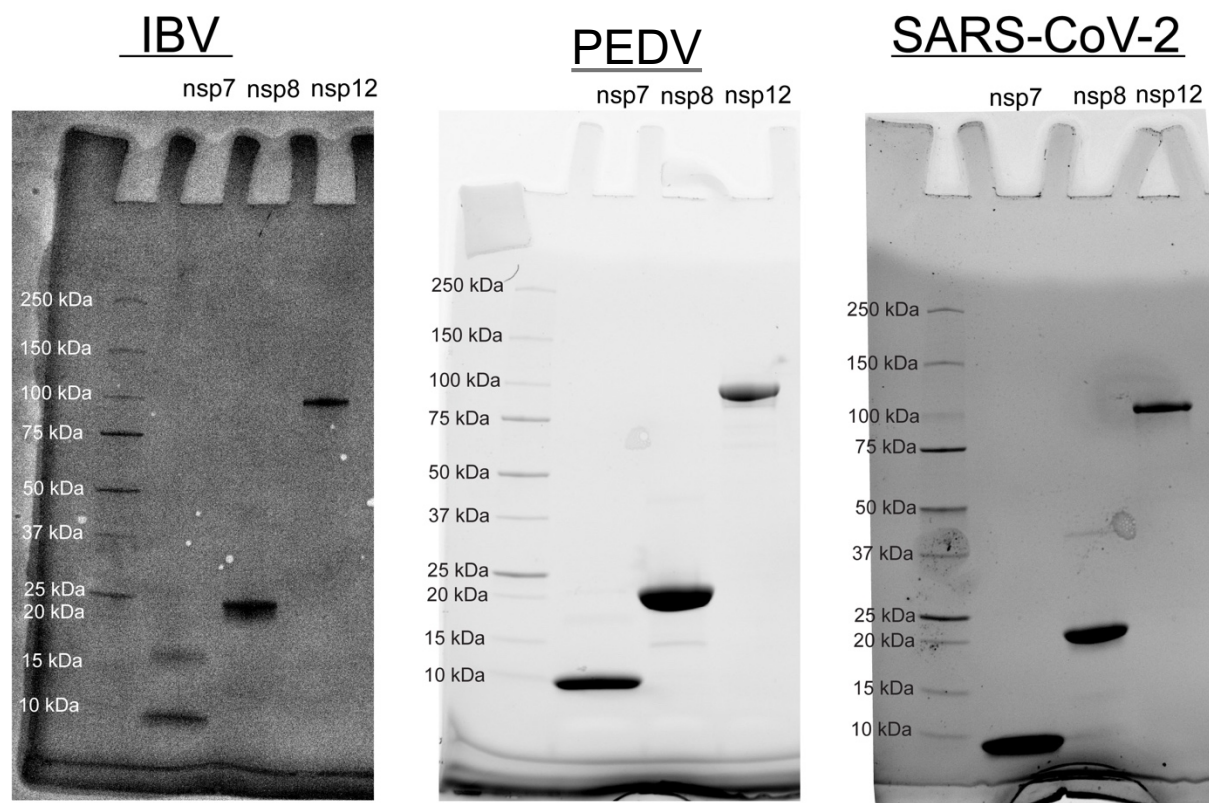


Figure S1 SDS-PAGE analysis of viral RTC proteins: Purified recombinant viral proteins used for activity assays and structure determination analyzed via SDS-PAGE and visualized using UV fluorescence. Each gel ran with ladder on far-left lane with molecular weights labeled. Expected MW of each protein is as follows: nsp7 – ~9 kDa, nsp8 – ~22 kDa, nsp12- ~108 kDa.

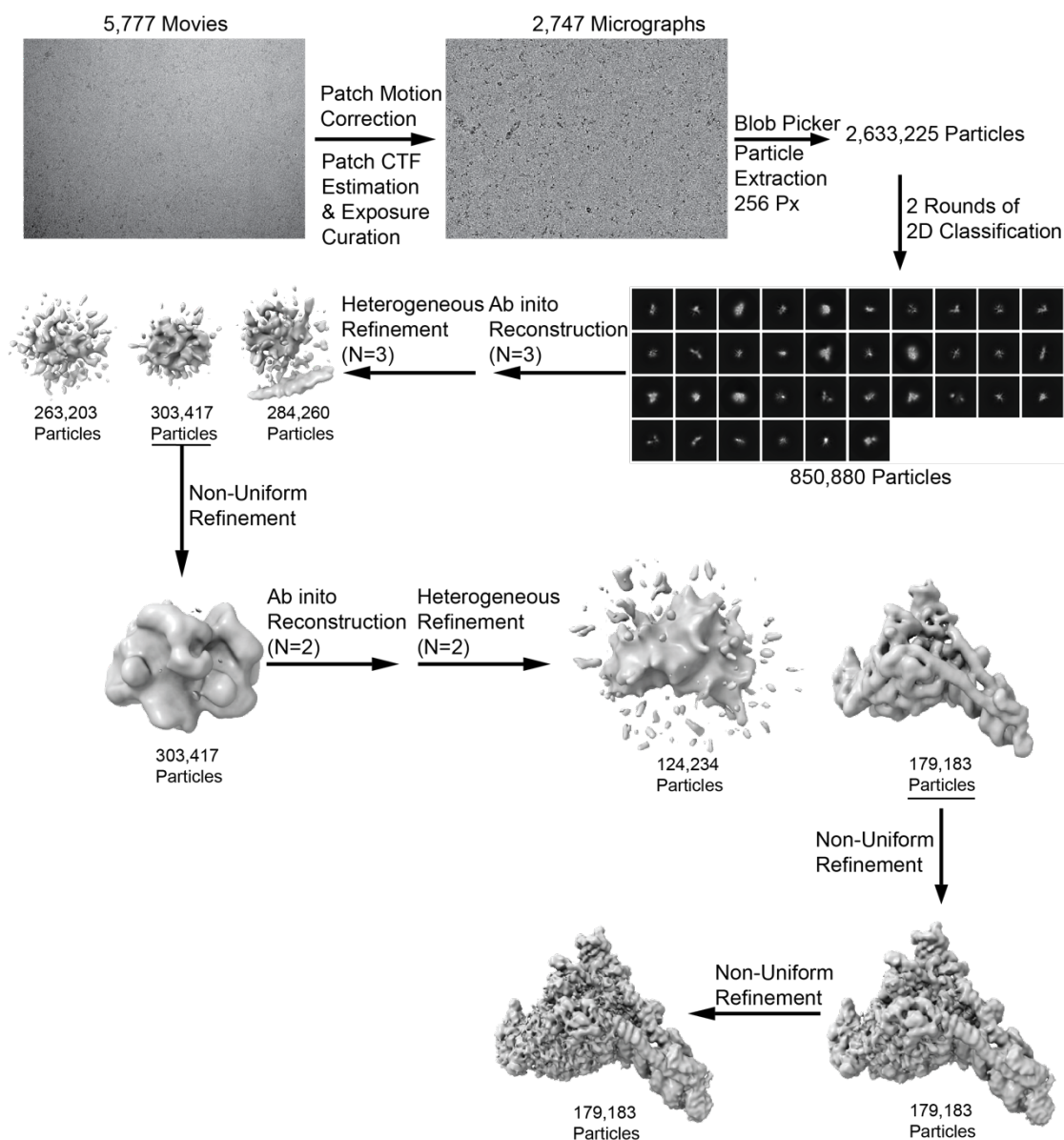


Figure S2 Cryo-EM data processing pipeline: Workflow for EM data processing using Cryo-Sparc. 5,777 movies were processed to micrographs and subsequently to individual particles which were filtered through two rounds of 2D classification. Ab initio reconstruction followed by heterogenous refinement was done twice to finalize on data set with 179,183 particles. Two rounds of non-uniform refinement were performed to give final refined 3.5 angstrom map.

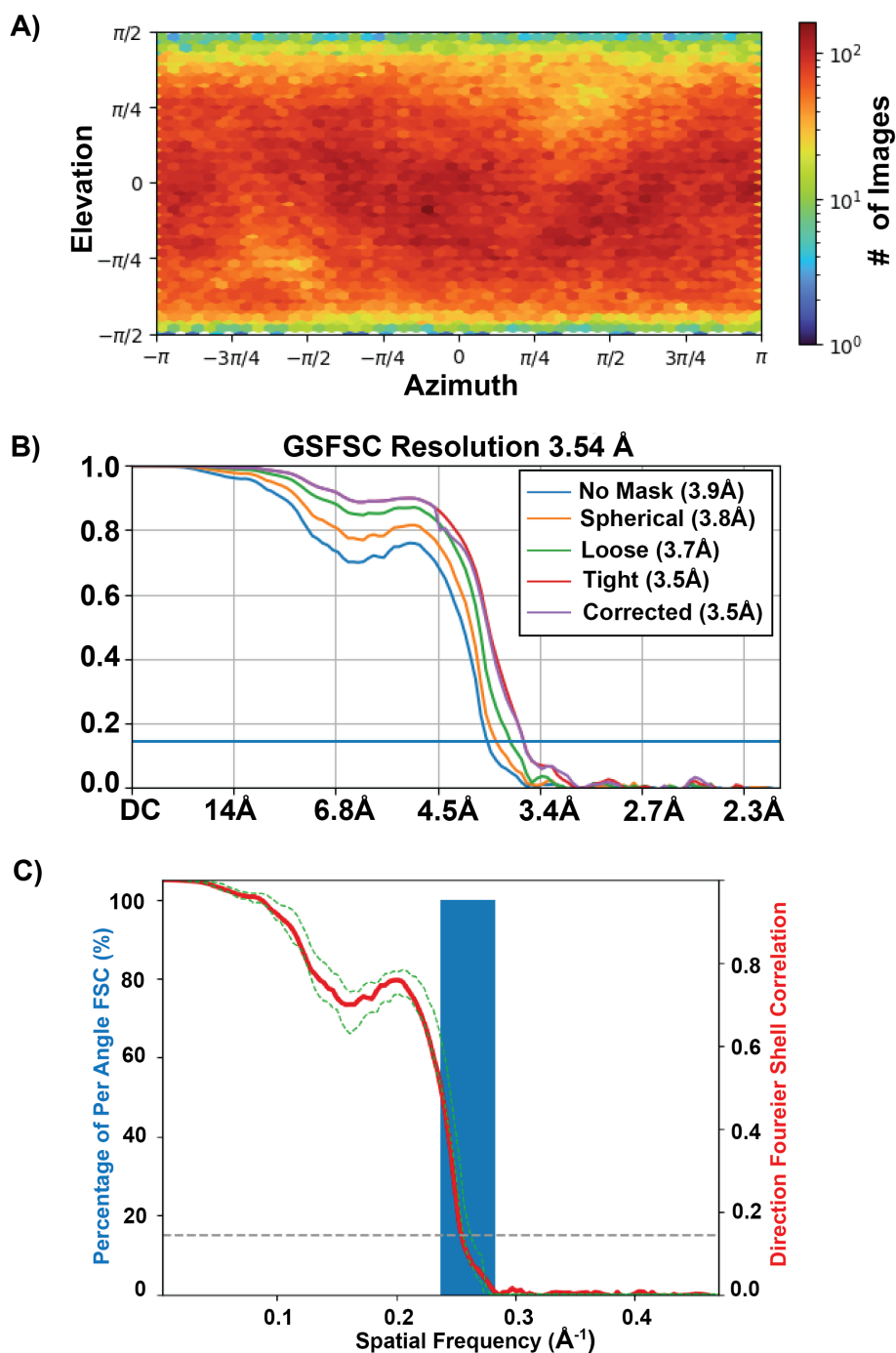


Figure S3 Cryo-EM data validation: A) Particle orientation distribution of for the final reconstructed EM map. B) Gold standard Fourier shell correlation plot from cryo-EM data. Blue line represents 0.143 cutoff. C) 3D Fourier shell correlation for the final refinement. Sphericity is 0.983 out of 1 and the unmasked global resolution is 3.95 angstrom.

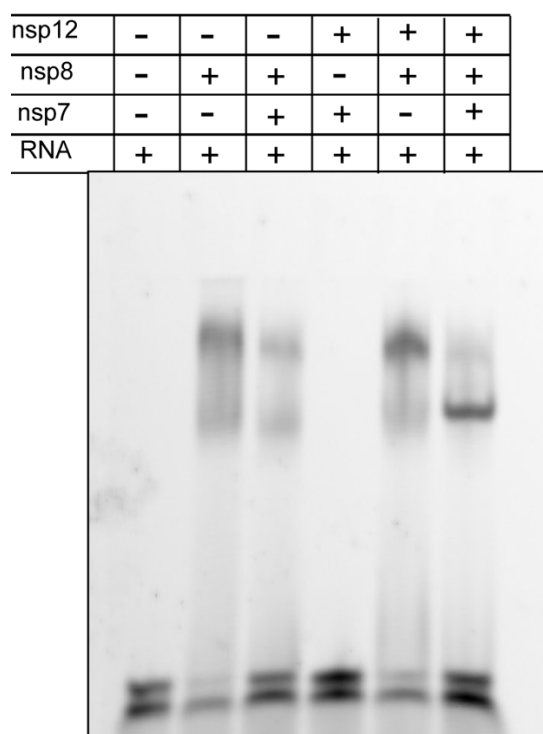


Figure S4 Native-PAGE IBV RTC RNA substrate affinity: Native PAGE gel analyzing IBV RTC binding to FAM tagged RNA duplex. Piecewise controls used to differentiate individual nsp vs whole complex interaction with RNA.

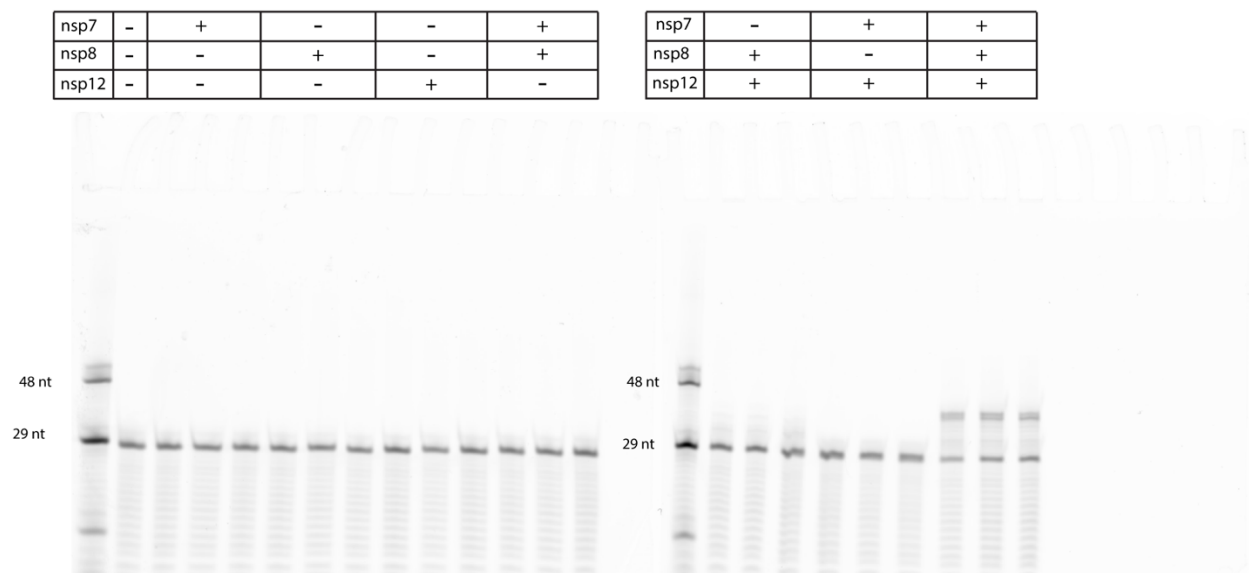
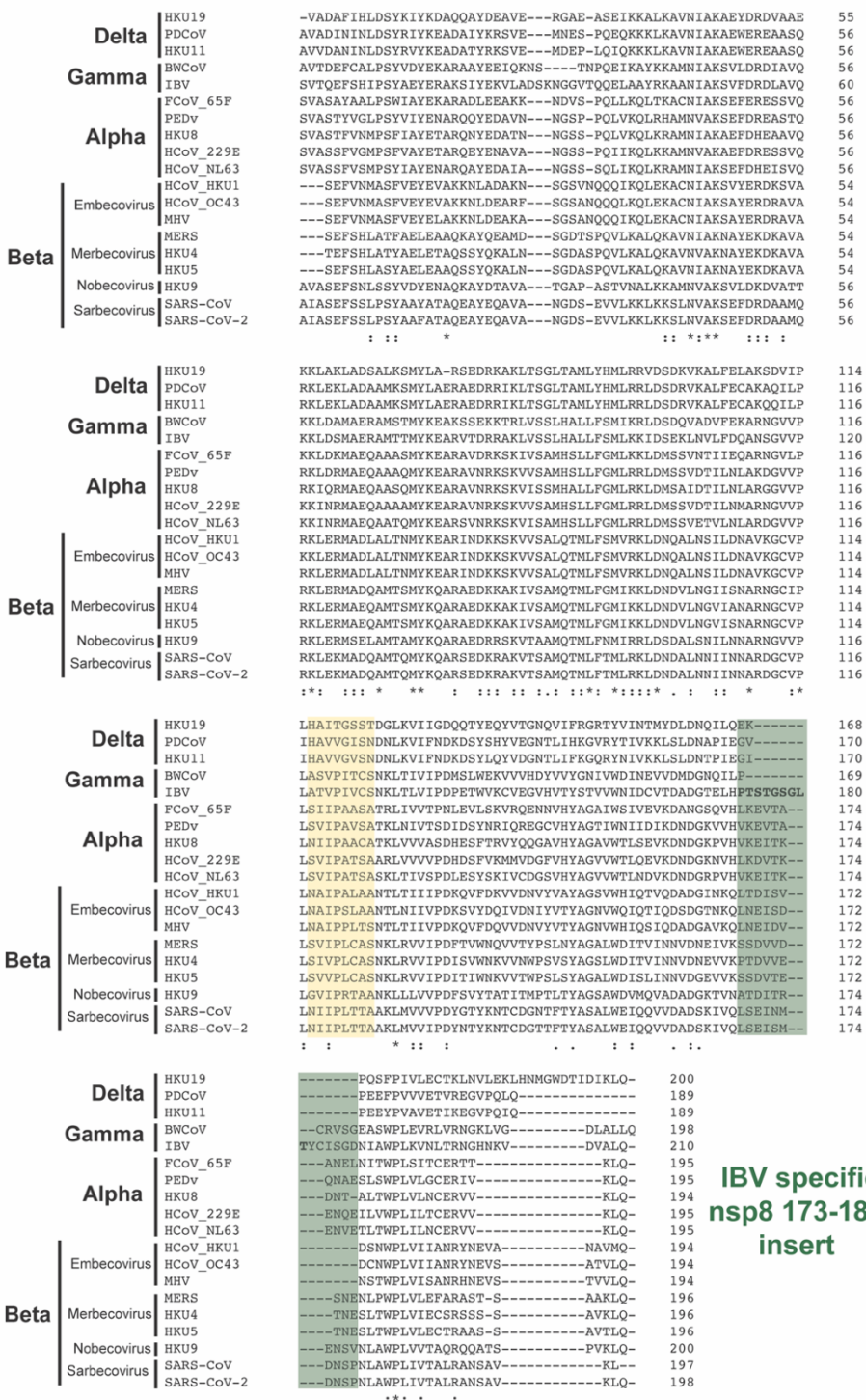


Figure S5 IBV RTC Activity Assay: Primer extension assay designed to assess activity of RTC based on ability to extend FAM tagged 29 nt RNA primer to the length of the 38 nt template. Triplicates ran with various combination of nsp7,8, and 12 mixed with RNA substrate. RNA analyzed on urea-PAGE gel.

Nsp8 Sequence Alignment



nsp8_γ variable
region 122-129

IBV specific
nsp8 173-181
insert

Figure S6 *Coronavirinae* nsp8 sequence alignment: Clustal omega alignment of coronavirus nsp8 proteins from alpha, beta, gamma, and delta genera. “*” notation means residue is fully conserved, “:” means residue is strongly conserved, and “.” Denotes a weakly conserved residue.

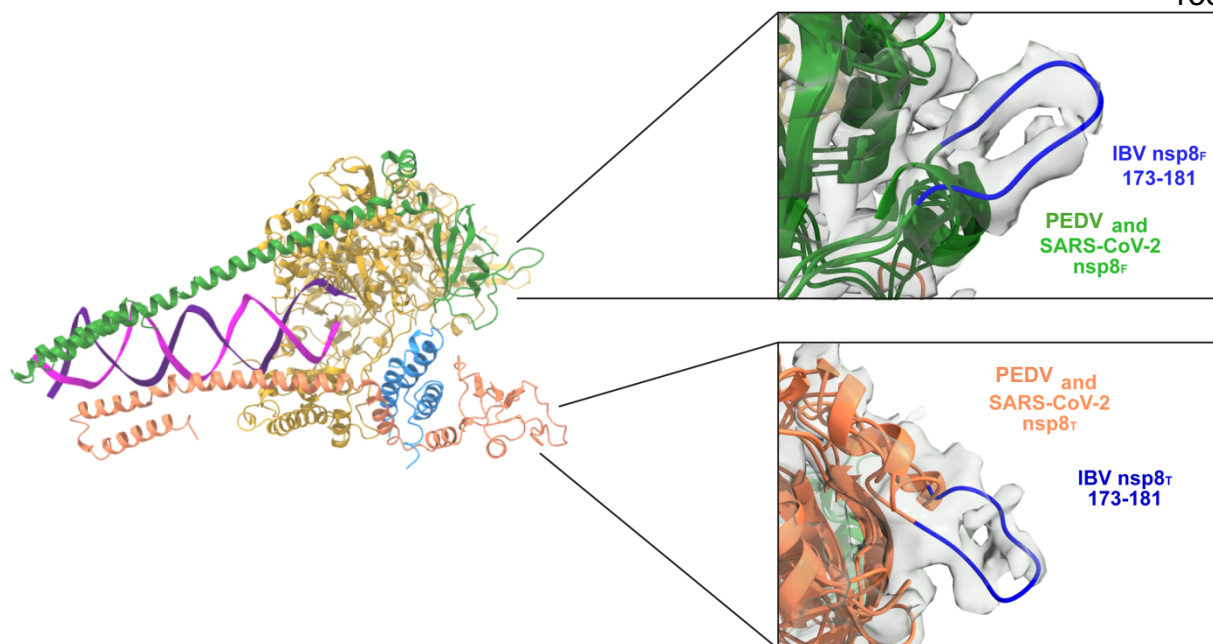


Figure S7 Structural view of IBV specific nsp8 173-181 insert: Cartoon model of the IBV complex (left) with zoomed in views of superimposed models of IBV, PEDV (8URB) and SARS-CoV-2 (6XEZ). Focused view highlights IBV nsp8 insert in both protomers found on RTC model with supporting electron density.

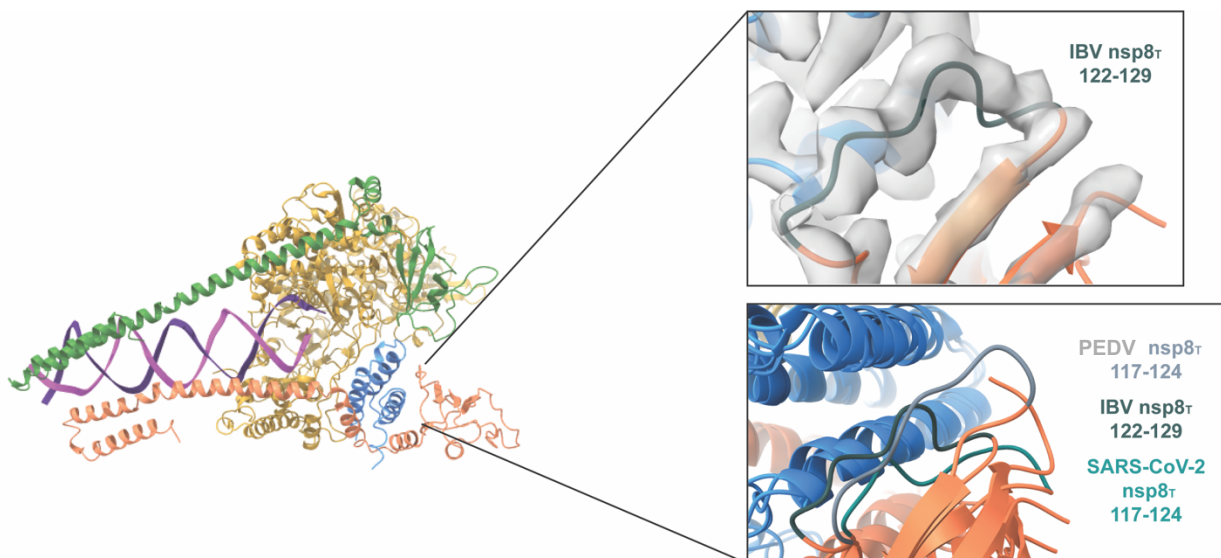


Figure S8 Structural view of nsp8_T 122-129 variable region: Cartoon model of the IBV complex (left) with zoomed in views of superimposed models of IBV, PEDV (8URB) and SARS-CoV-2 (6XEZ). Focused view highlights IBV nsp8_T 122-129 that has observed variability across genera. IBV model electron density shown to support conformation.

Nsp12 Sequence Alignment

Delta	HKU19	-LGYSSNQNNNSYLN RVK GSS-DARLEPCTSDNRPD V V V R A F N I Y N N --ATAGIFKSTKNN	56	
	HKU11	-----NSPYLN RV T GSS-GARLEPQQPGVT D A V K R A F V H N N --T T S G I F L S T K T N	49	
Alpha	PDCV	-----NSAYLN RV T GSS-DARLELPQGPQ T P D A V K R A F V H N D --T T S G I F L S T K S N	49	
	FCoV_65F	-----G T T V D Q S Y L N R V G S S -A A R L E P C N -G T D T P D H V S R A F D I Y N K --D V A C I G K F L K T N	52	
	HCoV_229E	-----S F D S S Y L N R V G S S -A A R L E P C N -G T D I D Y C V R A F D I Y N K --D A S F I G K N L K S N	50	
	HCoV_NL63	-----S V D I S Y L N R A R G S S -A A R L E P C N -G T D I D K C V R A F D I Y N K --R V S F L G K C L K M N	50	
Gamma	HKU8	-----S L D N N Y L N R V G S S -A A R L E P C N -G T E P E H V I R A F D I Y N K --D V A C I G K F V K V N	50	
	PEDV	-----S T D M A Y L N R V G S S -A A R L E X C N -G T D T Q H V I R A F D I Y N K --D V A C I G K F L K V N	50	
	BWCoV	-----S V P Q N Y L N R V G L S -E A R L R P C A S G L L P D V V K R A F D L Y N S --N T A G H Y A S L K H N	51	
	IBV	S V A V A S G F D K N Y L N R V G S S -E A R L I P L A N G C D P D V V K R A F D V C N K --E S A G M F Q N L K R N	57	
Beta	HCoV_OC43	-----S K D T N F L N R V G S V D A R L V P C A S G L S T D V Q L R A F D I Y N A --S V A G I G L H L K V N	52	
	Embecovirus	HCoV_HKU1	-----S K D L N F L N R V G T S V N A R L V P C A S G L S T D V Q L R A F D I C N T --N R A G I G L Y Y K V N	52
	MHV	-----S K D T N F L N R V G T S V N A R L V P C A S G L D T D V Q L R A F D I C N A --N R A G I G L Y Y K V N	52	
	MERS	-----S K D S N F L N R V G S I V N A R I E P C S S G L T D V V R A F D I C N V K A K V A G I G K Y Y K T N	54	
Beta	Merbecovirus	HKU4	-----S K D T N F L N R V G S S V N A R L E P C S S G L T D V V R A F D I C N F K A R V A G I G K Y Y K T N	54
	HKU5	-----S K D S N F L N R V G S I V N A R I E P C A S G L T D V V R A F D I C N Y K A K V A G I G K Y Y K T N	54	
	Nobecovirus	HKU9	-----A K D E C F L N R V G T S G V A R L V P L G S G V Q P D I V L R A F D I C N T --K V A G F G L H L K N N	52
	Sarbecovirus	SARS-CoV-2	-----S A D A Q S F L N R V C G V S -A A R L T P C G T G T S T D V V R A F D I Y N D --K V A G F A K F L K T N	52
	SARS-CoV	-----S A D A S T F L N R V C G V S -A A R L T P C G T G T S T D V V R A F D I Y N E --K V A G F A K F L K T N	52	

Delta	HKU19	CTRFKSTRP G S I I N K P V R T I E T F P V T K Q C T E N V F R A E E Q C Y N M L P K S I V S T D D K F S C V A Y	116	
	HKU11	C S R F K T T K N L P L P -N G S V E L Y P V S K Q C S Q Q V F E I E E T C Y N M F D S L K S T P E K P G V L A R	108	
Alpha	PDCV	C A R F K T T R S A L P L P -N G G S V E L Y P V T K Q C A A K V F E I E E C E Y N A L S T E L Y T D D T P G V L A K	108	
	FCoV_65F	C S R F R N L D -----K R D A Y Y V V K R C T K S V M D H E Q V C Y N D L K D -----S G A V A E	94	
	HCoV_229E	C V R F K N V D -----K D D A F Y I V K R C I K S V M D H E Q S M Y N L L K G -----C N A V A K	92	
	HCoV_NL63	C V R F K N A D -----L K D G Y F V I K R C T K S V M E H E Q S M Y N L L N F -----S G A L A E	92	
Gamma	HKU8	C V R F K N A D -----K H D A F Y V V K R C T K S V M E H E Q S I Y D A L K D -----C G A V S P	92	
	PEDV	C V R L K N L D -----K H D A F Y V V K R C T K S A M E H E Q S I Y S R L E K -----C G A V A E	92	
	BWCoV	C A R F Q E L D E N D -----E I D G F P V V K Q T P H M F E H E E K Y L D L K A -----D C V A V	96	
	IBV	C A R F Q E V R D T E D -----G N L E Y C D S Y P V V K Q T P P S N Y E H E K A C Y E D L K S -----E V T A D	106	
Beta	HCoV_OC43	C C R F Q R V D E N G D -----K L D Q F F V V K R T D L T I Y N R M K C Y E R V K D -----C K F V A E	98	
	Embecovirus	HCoV_HKU1	C C R F Q R I D D D G N -----K L D K F F V V K R T N L E V N Y N K E K T Y E L T F K S -----C G V V A E	98
	MHV	C C R F Q R V D D G N -----K L D K F F V V K R T N L E V N Y N K E K C Y E L T K E -----C G V V A E	98	
	MERS	T C R F V L D D Q G H -----L H D S Y F V V K R H T M E N Y E L E K H C Y D L L R D -----C D A V A P	100	
Beta	Merbecovirus	HKU4	T C R F V Q V D D E G H -----K L D S Y F I V K R H T M S N Y E L E K R C Y D L L K D -----C D A V A I	100
	HKU5	T C R F V E V D D E G H -----R L D S F F V V K R H T M E N Y E L E K R C Y D L V K D -----C D A V A V	100	
	Nobecovirus	HKU9	C C R Y Q E L A D A G N -----Q L D S Y F V V K R H T E S N Y L L E Q R C Y E K L K D -----C D V V A R	98
	Sarbecovirus	SARS-CoV-2	C C R F Q E K D E D D N -----L L D S Y F V V K R H T F S N Y Q H E E T I Y N L K D -----C P A V A K	98
	SARS-CoV	C C R F Q E K D E E G N -----L L D S Y F V V K R H T M S N Y Q H E E T I Y N L V K D -----C P A V A V	98	

IBV nsp12
67-72 insert

Delta	HKU19	H D F F K F D G ----V N N V V R R H L T K Y T L L D L V A L R H L S T S -Q E I I Q E I L T T M C G T S E E ---	168	
	HKU11	T E F F K F D K ----I P N V N R Q L T K Y T L L D L A Y A L R H L S T S -R D V I K E I L I T I C G T T E E ---	160	
Alpha	PDCV	T E F F K F D K ----I P N V N R Q L T K Y T L L D L A Y A L R H L S T S -K D V I Q E I L I T M C G T P E D ---	160	
	FCoV_65F	H D F F L Y K E G R C F G N V A R K D L T K Y T M D L C Y A I R N F D E K N C E V L K E I L V T L G A C N E S ---	151	
	HCoV_229E	H D F T T W H E G R T I Y G N V S R Q D L T K Y T M D L C F A L R N F D E K D C G V F E H I I L V T G C C N T D ---	149	
	HCoV_NL63	H D F F T W K D G R V I Y G N V S R H L T K Y T M D L V Y A M R N F D E Q N C D V L K E V L V L T G C C D N S ---	149	
Gamma	HKU8	H D F F W K D G R S V Y G N I A R H D L T K Y T M D L V A L R H F D E K N C E T L K E I L V T S G A C D S S ---	149	
	PEDV	H D F F T W K D G R A I Y G N V R K D L T E Y T M D L C Y A L R N F D E N N C D V L K S I L I K V G A C E E S ---	149	
	BWCoV	H D F F R F E -G ----M Y S I C R Q R L T K Y T M D L C Y A F R H F D P N D C D V L K E I L V V G C C E W D ---	149	
	IBV	H D F F V F N K N ----Y N I S R Q R L T K Y T M D F C Y A L R H F D P K D C E V L K E I L V T Y G C I E D Y H F K	163	
Beta	HCoV_OC43	H D F F T F D V G E S R V P H I V R K D L T K Y T M L D L C Y A L R H F D R N D C M L L C D I L S I Y A G C E Q S ---	155	
	Embecovirus	H D F F T F D I G S R V P H I V R N L S K Y T M L D L C Y A L R H F D R N D C S I L C E I L C E Y A D C K E S ---	155	
	MHV	H E F F T F D V G E S R V P H I V R K D L S K P T M L D L C Y A L R H F D R N D C S T L K E I L L Y A E C E E S ---	155	
	MERS	H D F F I P D V K V K T P H I V R Q L T E Y T M D L V Y A L R H F D Q N -S E V L K A I L V K Y G C C D V T ---	156	
Beta	Merbecovirus	H D F F I P D V D K T K T P H I V R Q S L T E Y T M D L V Y A L R H F D Q N N C E V L K S I L V K Y G C C E Q S ---	157	
	HKU4	H D F F I P D V D K V K T P H I V R Q L T E Y T M D L V Y A L R H F D Q N N C E V L K S I L V K Y G C C D A S ---	157	
	HKU5	H D F F K F N I E G V M T P H V S R E R L T K Y T M A D L V S L R H F D N N N C D T L K E I L L V T G C C T V D ---	155	
	Nobecovirus	HKU9	H D F F K F R I D G D M V P H I S R Q R L T K Y T M A D L V Y A L R H F D E G N C D T L K E I L V T Y N C C D D D ---	155
Beta	Sarbecovirus	SARS-CoV-2	H D F F K F R V D G D M V P H I S R Q R L T K Y T M A D L V Y A L R H F D E G N C D T L K E I L V T Y N C C D D D ---	155
	SARS-CoV		*** : *	

IBV nsp12
113-115
shortened loop

IBV nsp12
156-169 loop
insert

Delta	HKU19	W F --V D G W Y D P I E N P T F Y D E F H K L G S L I N N C V V M A N K F A D T C T K T V G L V G I L T A D N Q D L G G	226	
	HKU11	W F --G D S W F D P I E N P T F Y D E F H K L G S V L N R C V L N A N A F A K A C S E L G I V G I L T P D N Q D L G G	218	
Alpha	PDCV	W F --G E N W F D P I E N P S F Y K E F H K L G D I L N R C V L N A N K F A S A C I D A G L V G I L T P D N Q D L G G	218	
	FCoV_65F	-F F E N K W F D P V E N E A I H E V Y A L R G P I V A N A M L K C V A F C D A I V E K G Y I G I I T L D N Q D L N G	210	
	HCoV_229E	-Y F E M K W F D P I E N E D I H R V Y A A L G K V V A N A M L K C V A F C D E M V L K G V V G V L T L D N Q D L N G	208	
	HCoV_NL63	-Y F D S K G W Y D P V E N E D I H R V Y A S L G I V A R A M L K C V A L C D A M V A G V V G V L T L D N Q D L N G	208	
Gamma	HKU8	-Y F D N K W Y D P V E N E D I H R V Y A K L G C V V A N A M L K C V A L C D A M V A G V V G V L T L D N Q D L N G	208	
	PEDV	-Y F N K W F D P V E N E D I H R V Y A L L G T I V S R A M L K C V K F C D A M V E Q G I V G V L T L D N Q D L N G	208	
	BWCoV	-Y F D Q P W Y D P V E N P D W F S I G R L G P I F Q R A L K V A E F C D L M V E K G Y I G U V T L D N Q D L N G	208	
	IBV	W F E N K W Y D P I E N P K Y A M A K M G P I V R R A L L N A I E F G N L M V E K G Y V G V I T L D N Q D L N G	223	
Beta	HCoV_OC43	-Y F T K D W Y D P V E N P D I I N V Y K L G P I F N R A L V S A T E F A D K L V E V G L V G V L T L D N Q D L N G	214	
	Embecovirus	HCoV_HKU1	-Y F S K D W Y D P V E N P D I I N I Y K L G P I F N R A L L N T I V A D T L V E V G L V G V L T L D N Q D L Y G	214
	MHV	-Y F Q K D W Y D P V E N P D I I N V Y K L G P I F N R A L L N T A K F A D A L V E A G L V G V L T L D N Q D L Y G	214	
	MERS	-Y F E N K W F D P V E N P S I G V Y H K L G E R V R Q A I L N T V K F C D H M V K A G L V G V L T L D N Q D L N G	215	
Beta	Merbecovirus	HKU4	-Y F D N K L W F D P V E N P S I G V Y H K L G E R I R Q A M L N T V K M C D H M V K S G L V G V L T L D N Q D L N G	216
	HKU5	-Y F D N K L W F D P V E N P V I S V Y H K L G E R I R Q A V L N T V K F C D Q M V K S G L V G V L T L D N Q D L N G	216	
	Nobecovirus	HKU9	-Y F R K D W Y D P V E N P D I R V Y H K L G E R V R Q A L K T V Q F C D A M R N A G I V G V L T L D N Q D L N G	214
	Sarbecovirus	SARS-CoV-2	-Y F N K D W Y D P V E N P D I L R V Y A N L G E R V R Q A L K T V Q F C D A M R N A G I V G V L T L D N Q D L N G	214
	SARS-CoV	-Y F N K D W Y D P V E N P D I L R V Y A N L G E R V R Q S L L K T V Q F C D A M R A D I V G V L T L D N Q D L N G	214	

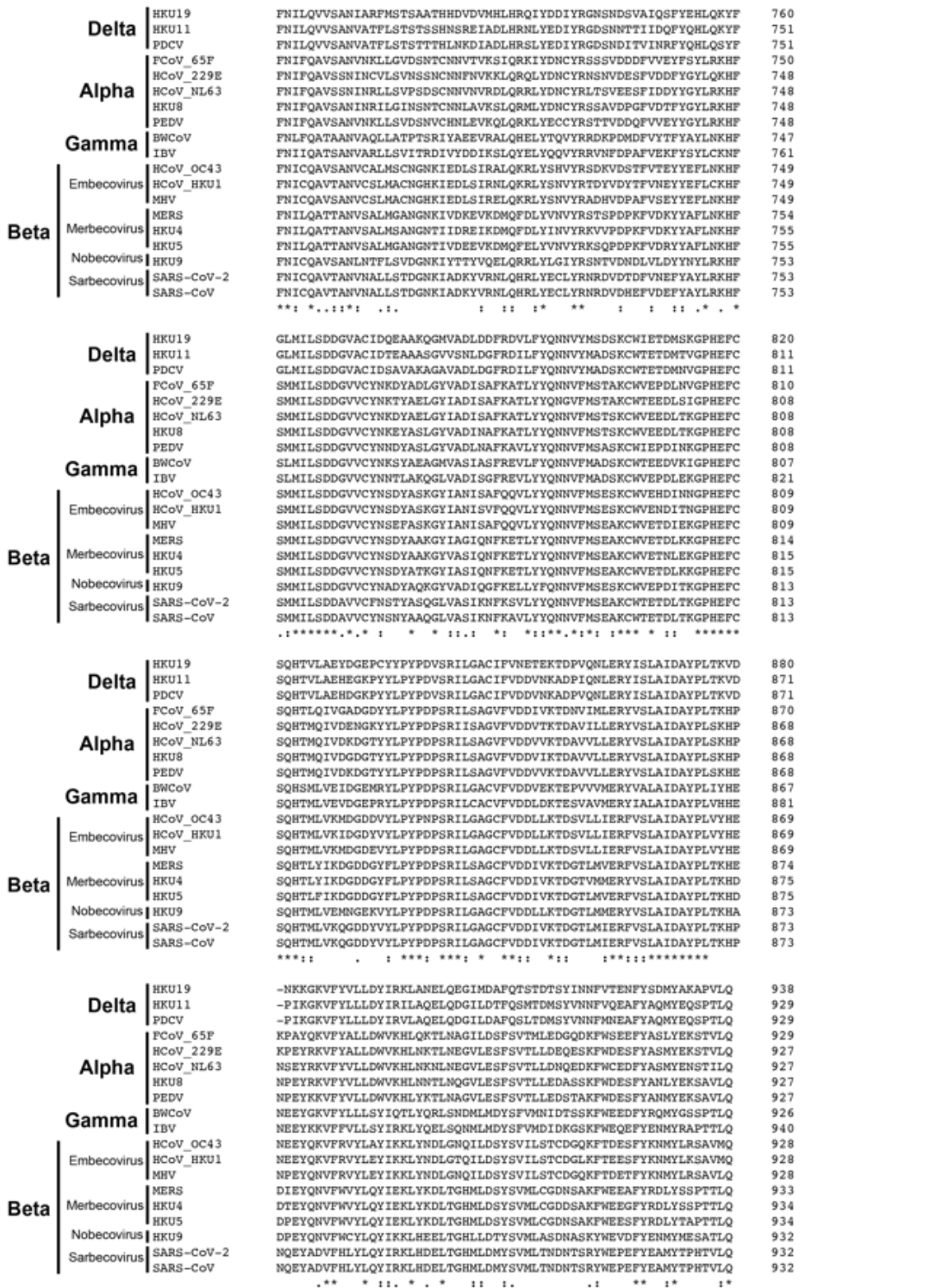


Figure S9 *Coronavirinae* nsp12 sequence alignment: Clustal omega alignment of coronavirus nsp12 proteins from alpha, beta, gamma, and delta genera. “*” notation means residue is fully conserved, “:” means residue is strongly conserved, and “.” Denotes a weakly conserved residue.

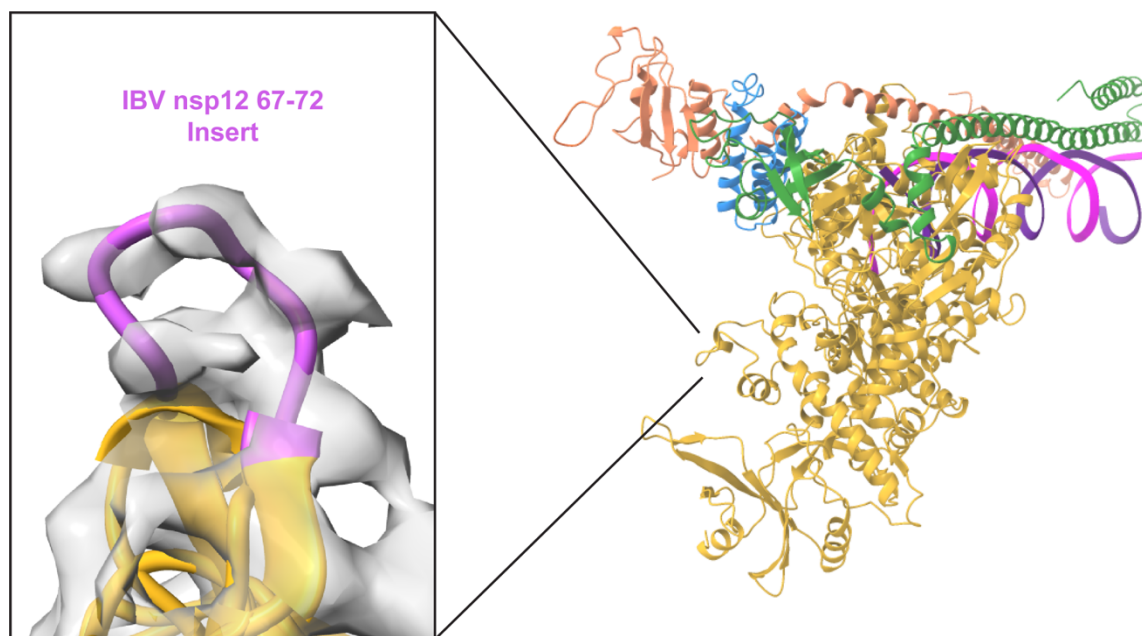


Figure S10 Structural view of IBV nsp12 67-72 insertion: Cartoon model of the IBV complex (right) with zoomed in view of superimposed models of IBV, PEDV (8URB) and SARS-CoV-2 (6XEZ). Focused view on IBV nsp12 67-72 insertion with electron density to support inserted region conformation.

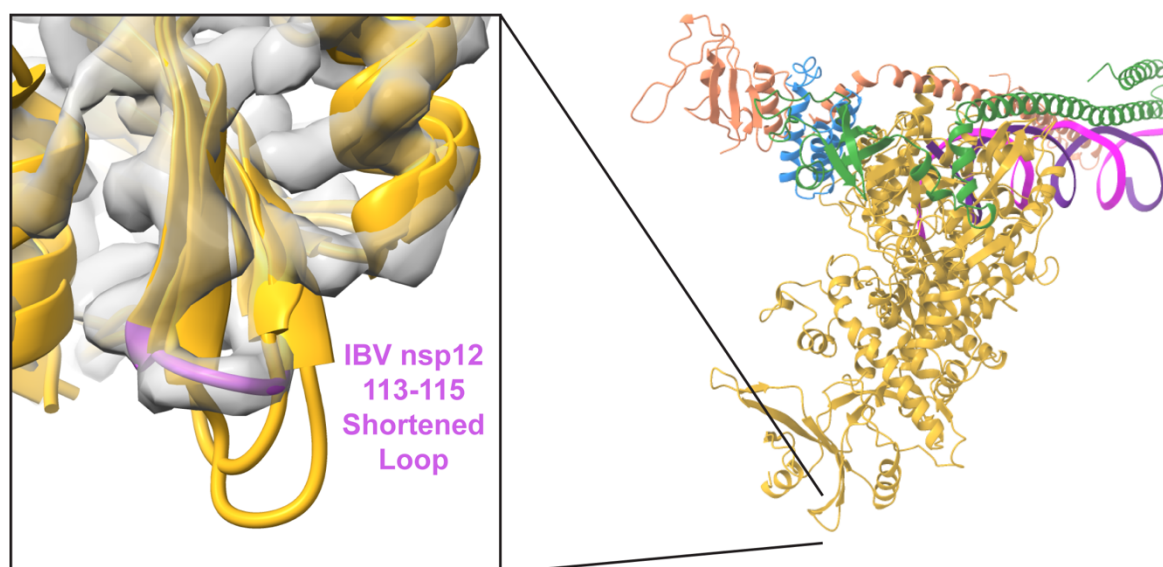


Figure S11 Structural view of IBV nsp12 113-115 shortened loop: Cartoon model of the IBV complex (right) with zoomed in view of superimposed models of IBV, PEDV (8URB) and SARS-CoV-2 (6XEZ). Focused view on IBV nsp12 shortened loop with electron density to support loop conformation.

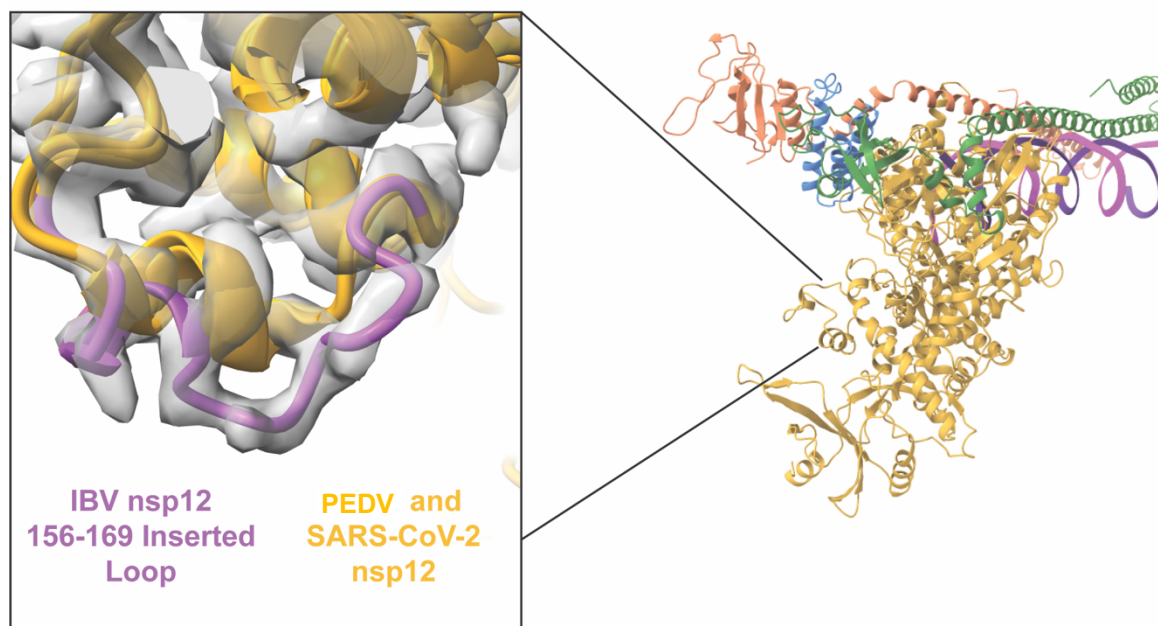
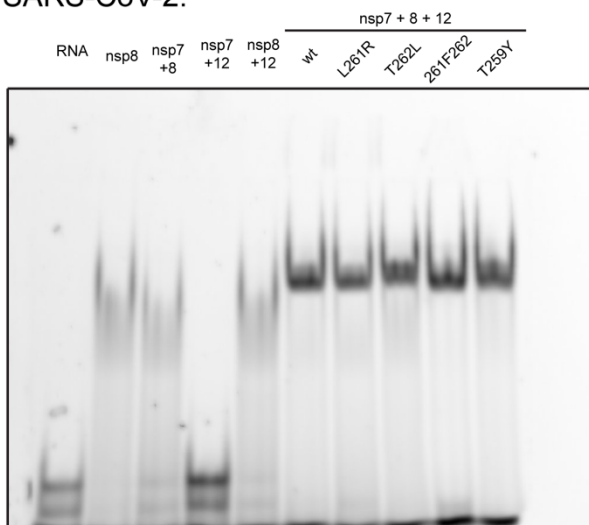


Figure S12 Structural view of IBV nsp12 156-169 insertion: Cartoon model of the IBV complex (right) with zoomed in view of superimposed models of IBV, PEDV (8URB) and SARS-CoV-2 (6XEZ). Focused view on IBV nsp12 156-169 which has a 4 amino acid insertion. IBV model electron density shown to support altered loop conformation.

SARS-CoV-2:



IBV:

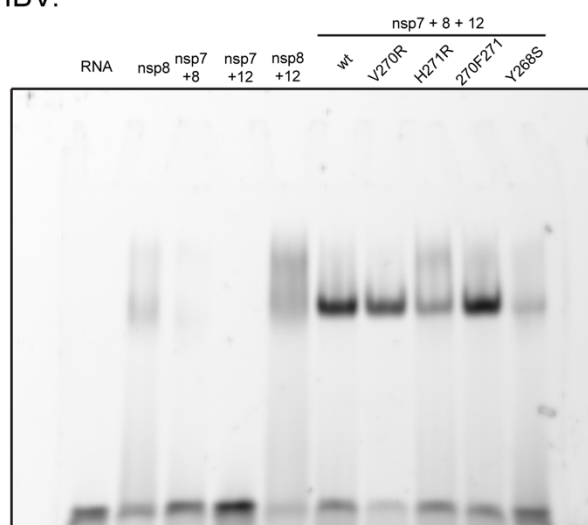


Figure S13 Mutant nsp12 RTC RNA substrate affinity: Native PAGE gel analyzing RTC binding to FAM tagged RNA duplex. Separate gels ran comparing wild type to mutant for each SARS-CoV-2, and IBV.

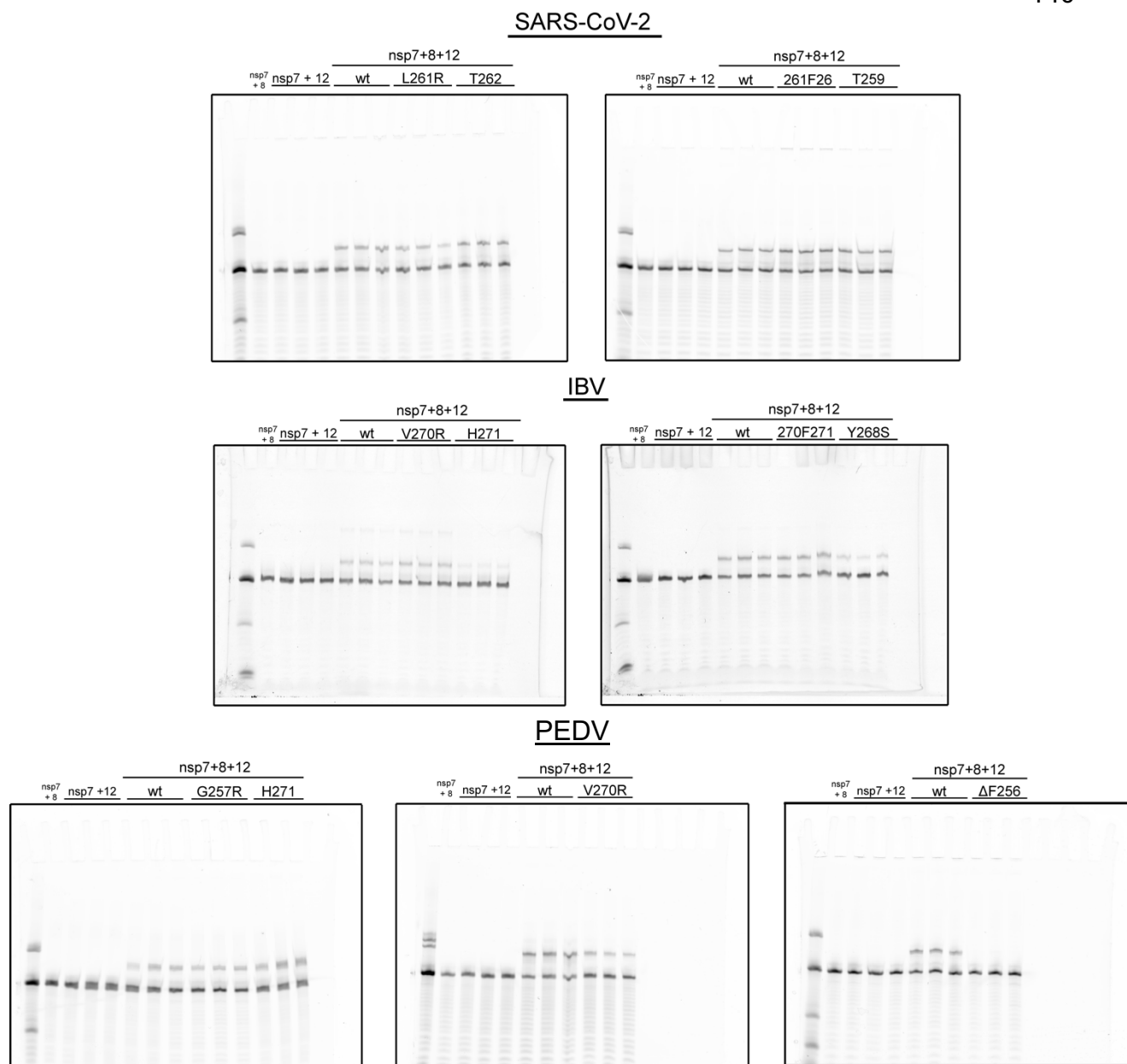


Figure S14 Mutant nsp12 primer extension assays: Primer extension assay assessing RTC activity with different mutant nsp12s for SARS-CoV-2, IBV and PEDV. Each gel contains the same controls of nsp7+8, nsp7+12 and wild type nsp7+8+12 performed in triplicate.

**Chapter 4: The coronavirus exonuclease and polymerase interaction is
dependent on RNA backtracking**

Authors: Thomas K. Anderson¹, Kennan J. Chojnacki¹, Peter J. Hoferle¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A. produced the data for Figures 2A, 3D, and 4

K.J.C. produced the data for Figures 2B and C, and 3A, B, and C

T.K.A. and R.N.K. developed this project

T.K.A. wrote this chapter with input and editing by R.N.K.

Abstract:

Coronaviruses (CoVs) are a subfamily of (+) single-stranded RNA (ssRNA) viruses that infect humans and animals causing a wide range of disease severities. In 2019, SARS-CoV-2 emerged from animal reservoirs and is the causative agent of the COVID-19 pandemic. At approximately 30 kb, CoV genomes are substantially larger than most other RNA viruses. CoVs maintain their large RNA genomes using the virally encoded exonuclease, nsp14. Coronaviruses, and other closely related viruses, are the only known RNA viruses to proofread during viral RNA replication. While nsp14 is known to mediate proofreading, how it does so has remained elusive. Among the possible mechanisms of CoV proofreading are nsp14's association with the viral polymerase complex or its action on RNA independently. In our present work, we determine the requirements for the binding of nsp14 to the polymerase complex and narrow down the possible interaction sites. Our work provides critical insight into the unique mechanism of substrate engagement by the CoV RNA proofreading complex.

Introduction:

Coronaviruses (CoVs) are a subfamily of viruses within the order *Nidovirales* that have a propensity to crossover from animal reservoirs into humans causing epidemics or pandemics (1,2). SARS-CoV emerged in 2002 causing >8,000 cases and >800 deaths (3,4). Since 2012, MERS-CoV has continued to crossover from animal reservoirs and cause sporadic outbreaks with mortality rates as high as 35% (5-7). In 2019, SARS-CoV-2 emerged and is the causative agent of the ongoing COVID-19 pandemic which has caused over 7 million deaths globally (8,9). In addition, there are recent reports of other zoonotic CoVs infecting humans that have yet to cause large outbreaks, including the *DeltaCoV* porcine delta CoV and the *AlphaCoV* CCoV-HuPn-2018 (10-13).

CoVs have (+) single-stranded RNA (ssRNA) genomes that average in size around 30 kb (14,15). CoV genomes have 5' caps and 3' poly-adenosine tails (14,15). The 5' two-thirds of the genome contains two large open reading frames (ORFs) that encode for 16 non-structural proteins (nsps) (16-18). The majority of the nsps are believed to assemble into the viral replication-transcription complex (RTC) that synthesizes and modifies viral RNA products (19,20). The central nsp of the RTC is the viral RNA-dependent RNA polymerase (RdRP) nsp12 (21,22). Two protein replication factors, nsp7 and nsp8, bind nsp12 to form the viral core-RTC, a highly processive RNA polymerase complex with nucleotide addition rates as high as 170 nt/second (21-23).

Structures of the CoV core-RTC revealed that each complex binds two nsp8s; these are delineated by whether they bind the fingers (nsp8_F) or thumb (nsp8_T) domain of nsp12 (24-26). Nsp7 and nsp8_T form a heterodimer on the core-RTC (24). Each nsp8 has a long helical N-terminal extension that binds dsRNA exiting the RdRP active site

(25). These interactions have been predicted to be critical for processive RNA elongation. In support of this, it has been previously shown that mutating positively charged residues in these extensions' blocks replication (21,25). More recent work on SARS-CoV-2 and the *AlphaCoV* PEDV RTCs has demonstrated that both helical extensions are not required for RNA synthesis *in vitro*, indicating that they may have other roles in replication (22,27). Structures of the SARS-CoV-2 RTC bound to the viral helicase, nsp13, revealed that two helicase protomers bind one RTC (**Fig 1**) (28,29). Each nsp13 is scaffolded to the RTC by one of the nsp8s (28,29). The nsp13 scaffolded by nsp8_T binds the 5' end of template RNA (**Fig 1**) (28,29). This observation led to the prediction that this nsp13 could drive RTC backtracking, as the direction of the helicase (5'-3') counters that of the polymerase on the template RNA (3'-5') (**Fig 1**) (28). Further work demonstrated that the SARS-CoV-2 RTC can bind a frayed RNA duplex with five nucleotide mismatches of ssRNA at the primer 3' end (30). Assembly with this RNA duplex formed a backtracked RTC, with the 3' mismatches backtracked through nsp12's NTP channel (**Fig 1**) (30). Although unlikely to occur naturally, multiple consecutive mismatches are used to model a polymerase in a backtracked state. Similar designs have been used in structural studies of other polymerases (31,32). Additional biochemical experiments validated nsp13's ability to drive RTC backtracking (30). Currently, the signal, purpose, and regulation of backtracking are unknown. One hypothesis is that backtracking occurs after nucleotide mismatch introduction; mismatches are known to produce frayed duplexes that promote backtracking in DNA and RNA polymerases (30,32,33).

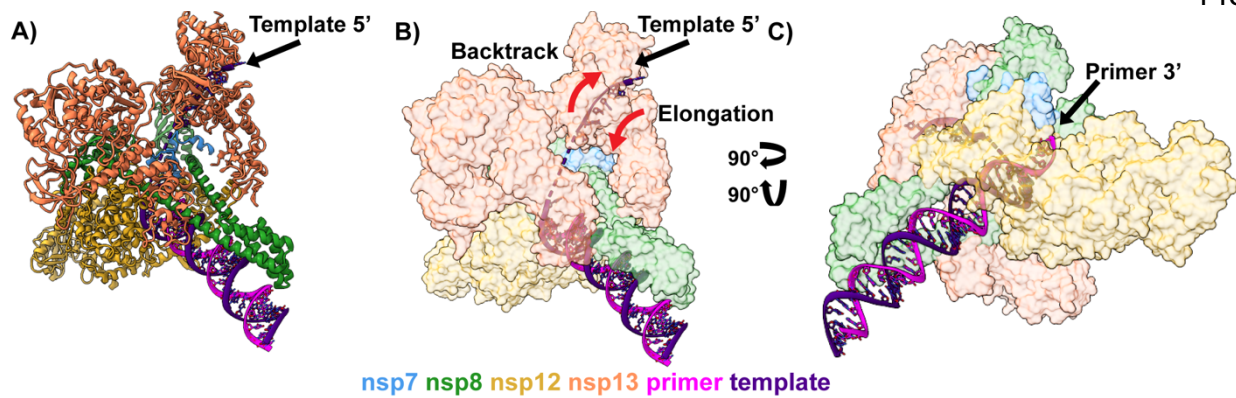


Figure 1. RTC backtracking: Structure of the SARS-CoV-2 RTC (PDB: 7KRO) bound to the viral helicase, nsp13, in a backtracked state. **A and B)** The nsp13 protomer bound to nsp8_T binds the 5' end of template RNA. Red arrows indicate the direction of RNA during elongation or backtracking. **C)** The 3' end of nascent RNA extrudes out of nsp12's NTP channel during backtracking. In **B** and **C**, proteins are shown as transparent surfaces and RNA as a ribbon model to emphasize the location of the RNA.

RNA viruses typically have much smaller genomes (often <20 kb) when compared to the large range of genome sizes across DNA viruses (up to 2 Mbp) (34). RNA genomes are believed to be restricted in size due to the low fidelity of virally encoded RdRPs (35,36). The low fidelity of RdRPs is a tradeoff for fast replication rates that allow RNA viruses to produce many genome copies during an infection (37). The population of similar but genetically distinct genomes made during an infection is referred to as a quasispecies (36,38). The mutagenic rates that produce quasispecies cause RNA viruses to exist near an error threshold, a theoretical limit to the number of mutations a virus population can sustain (39,40). Surpassing this threshold results in unproductive replication and virus population extinction, an event termed error catastrophe (39-41). Existing near this error threshold has been coopted in the design of antiviral drugs that increase mutagenic rates to induce error catastrophe (41,42).

The (+) ssRNA virus order *Nidovirales* contains several viral families with genomes >20 kb: *Roniviridae*, *Mesoniviridae*, and *Coronaviridae* (1,43). The largest known RNA virus is a nidovirus that has a genome >40 kb (44). Having such large genomes presents a challenge for RNA viruses, as larger genomes increase the chance for mutations, and further, error catastrophe to occur. Of the *Nidoviruses*, the CoV subfamily (within *Coronaviridae*) is the most characterized. How CoVs replicate their large RNA genomes without error catastrophe was puzzling until a 3'-5' exoribonuclease was identified in the CoV nsp14 during a genome bioinformatics analysis (4,45,46). Nsp14 belongs to the DE(E/D)D superfamily of exonucleases, defined by active site residues involved in catalysis (4,45). This superfamily includes exonucleases involved in DNA proofreading in prokaryotes and eukaryotes, such as the Klenow fragment (47,48). Identification of nsp14

as an exonuclease provided the first suggestion that CoVs may proofread during replication, potentially providing a mechanism to avoid error catastrophe (46). In support of this, other families in *Nidovirales* with genomes <20 kb, such as *Arteriviridae*, share many homologous proteins with CoVs, but lack a homologue for nsp14 (1).

Nsp14's 3'-5' exonuclease activity has been validated *in vitro* and it has been shown that the cofactor nsp10 is required for optimal activity (45,49). Reports on the exonuclease complexes' (nsp10+nsp14) substrate preference for ssRNA or dsRNA have varied, indicating that the activity differs depending on recombinant purification protocols and assays used to test activity (45,50,51). Nsp14's role as the proofreading enzyme has been validated using reverse genetics experiments. Exonuclease knockout of *BetaCoV* MERS-CoV, and SARS-CoV-2 fails to produce progeny virus (52). Knockout of activity in the *AlphaCoV* 229E resulted in decreased viral RNA synthesis and no progeny virus production (45). However, knockout of exonuclease activity in the *BetaCoV* SARS-CoV and MHV produces viable viruses with mutation rates 15-20 times that of wildtype viruses (53,54). CoVs' ability to proofread makes them naturally resistant to commonly used nucleoside analogue antiviral drugs such as ribavirin and 5-fluorouracil (55,56). Mutations to either knockout the exonuclease activity or nsp14's interaction with nsp10 can render CoVs (SARS-CoV and MHV) susceptible to these drugs (56,57).

There are multiple ways nsp14 could mediate proofreading, including mechanisms that involve nsp14 associating with the RTC or binding mismatched RNA directly without the RTC present (46,50). In support of the former, nsp14 has been shown to bind nsp8 and the RTC (21,58,59). The initial work that identified nsp14's interaction with the RTC excluded nsp10 from complex assembly assays, so the role of nsp10 in polymerase and

nsp14 interactions is currently unknown (21). A structure of the SARS-CoV-2 RTC with nsp10+nsp14 bound has been solved, but to stably assemble the complex an nsp9-nsp10 fusion protein was used (60). While interesting, the validity of this fusion complex structure has not been examined further.

At present, the data in the field of CoV replication yields contradictory results and theories as to the mechanism of CoV proofreading. While it is known that nsp14 is the proofreading enzyme, its interactions, or lack thereof, with the RTC remain disputed. To address this gap, we sought to determine if the SARS-CoV-2 exonuclease complex (nsp10+nsp14) interacts with the core-RTC (nsp7+nsp8+nsp12). Using different RNA substrates and biochemical assays, we have identified an RNA substrate requirement for complex protein interactions and narrowed down the potential site of exonuclease-RTC interaction. Better understanding the mechanisms by which CoVs replicate and proofread RNA will allow for the development of more effective antivirals that avoid or block proofreading.

Methods:

Expression Constructs:

SARS-CoV-2 nsp genes were codon optimized and synthesized from protein sequences associated with GenBank UHD90671.1 (Genscript). Nsp7, nsp8, nsp10, and nsp14 were cloned into pET46 expression vectors with N-terminal hexa-histidine tags, and enterokinase and TEV protease cleavage sites. Nsp12 was cloned into a pFastBac vector with C-terminal TEV protease site and double Strep-II tags. Mutant nsp12 and nsp14 expression plasmids were produced using lightning site-directed mutagenesis. The

sequences of all open reading frames within plasmids were confirmed using Sanger sequencing.

Recombinant Protein Expression:

SARS-CoV-2 nsp7, nsp8, and nsp10 were expressed in Rosetta 2pLysS *Escherichia coli* (*E. coli*) cells (Novagen). Cultures were grown at 37°C until they reached an OD₆₀₀ of 0.6-0.8 at which point they were induced by addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 500 μM. After shaking for 16 hours at 16°C cells were harvested by centrifugation and resuspended in Ni-Wash Buffer (10 mM Tris-Cl, pH 8, 300 mM NaCl, 30 mM imidazole, and 2 mM dithiothreitol (DTT)). Resuspended cells were lysed via microfluidization (Microfluidics), and lysates cleared by centrifugation followed by 0.45 μm filtration. Proteins were purified using Ni-NTA agarose beads (Qiagen), eluting with wash buffer containing 300 mM imidazole. Eluted protein was dialyzed and digested overnight with 1% TEV protease (w/w) at 4°C in dialysis buffer (10 mM Tris-Cl pH 8, 300 mM NaCl, 2 mM DTT). Undigested protein was removed by flowing samples over Ni-NTA agarose beads (Qiagen) and digested protein was further purified using a Superdex 200 Increase 10/300 GL column (Cytiva) with SEC buffer (25 mM Tris-Cl, pH 8, 300 mM NaCl, 2 mM DTT). Fractions containing the protein of interest were pooled and concentrated using ultrafiltration. After concentration, proteins were aliquoted, flash frozen in liquid nitrogen, and stored at -80°C. Replication factor proteins (nsp7, nsp8, and nsp10) were prepared in either 10 mM Tris-pH 8 or 25 mM HEPES-pH 7.5 buffers with all other salts and additives the same. The yield, purity and activity of proteins did not change whether they were prepared in Tris or HEPES buffers. Typical protein yields for a 1 L prep ranged from 10 – 40 mg.

SARS-CoV-2 nsp14 (wildtype or mutant) expression and purification followed an identical protocol to that of the replication factor proteins except for the following details. Nsp14 was expressed in C41 (DE3) *E. coli* cells (BioSearch Technologies, Lucigen). *E. coli* cell pellets were resuspended in nsp14 wash buffer (25 mM HEPES, pH 7.5, 300 mM NaCl, 30 mM imidazole, 2 mM MgCl₂, and 2 mM DTT). Protein was treated with 1% TEV protease (w/w) overnight at 4°C in dialysis buffer (25 mM HEPES, pH 7.5, 300 mM NaCl, 2 mM DTT). Typical yield for a 1 L prep was 1 mg of protein.

The pFastBac plasmid containing the SARS-CoV-2 nsp12 gene (wildtype or mutant) was used to transform DH10Bac *E. coli* cells (ThermoFisher Scientific) and prepare recombinant bacmids containing the nsp12 gene. Bacmids were transfected into Sf9 cells (Expression Systems) with CellFectin II (Life Technologies) to produce recombinant baculoviruses which were then amplified two times in Sf9 cells. The amplified baculovirus stock was used to infect 1 L of Sf21 insect cells (Expression Systems). After two days of shaking at 27°C, cell pellets were harvested by centrifugation and resuspended in nsp12 wash buffer (25 mM HEPES, pH 7.4, 300 mM NaCl, 1 mM MgCl₂, and 2 mM DTT) with an added 143 µL of BioLock (per 1 L of culture). Cells were lysed by microfluidization (Microfluidics), and lysates were cleared by centrifugation followed by 0.45 µm filtration. Nsp12 was then purified using Streptactin superflow agarose (IBA) beads and eluted with nsp12 wash buffer supplemented with 2.5 mM desthiobiotin. Nsp12 was further purified via SEC using a Superdex 200 Increase 10/300 GL column (Cytiva) in nsp12 SEC buffer (25 mM HEPES pH 7.4, 300 mM NaCl, 100 µM MgCl₂, and 2 mM tris(2-carboxyethyl)phosphine (TCEP)). Fractions with nsp12 were pooled and concentrated using ultrafiltration. After concentration, proteins were aliquoted,

flash frozen in liquid nitrogen and stored at -80°C . Typical yield for a 1 L culture was 1-5 mg of protein.

Sequence and Annealing of RNA Substrates:

RNA oligos were purchased from Integrated DNA Technologies (IDT). Primer RNAs were purchased with 5' fluorescein (6-FAM) tags to monitor the RNA by gel electrophoresis. Some primer RNAs (P RNA) contain a "*" which denotes a thiophosphate that was initially used to reduce potential RNA degradation though the use of nsp14 exonuclease knockouts made this addition unnecessary. To anneal dsRNA, template RNA (T RNA) was held at 20% excess to primer RNA. RNA was annealed in annealing buffer (2.5 mM HEPES, pH 7.4, 2.5 mM KCl, and 0.5 mM MgCl_2). After mixing, samples were heated at 95°C for 5 minutes, then allowed to slowly cool for 1.25 hours back down to 25°C . Once annealed, samples were stored at -20°C .

RNA:	Sequence:
P0	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU-3'
P4	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGA -3'
P7	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU -3'
P8	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU *C-3'
P10	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU *CCC-3'
P14	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU *CCCUUCU-3'
P15	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU CCCUUCUA-3'
P16	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU CCCUUCUAC-3'
P18	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU CCGAGCU *CCCUUCUACCU-3'
T-Backtrack	3' -GUAAGAGGAUUCUUCGAUAAUUUUAGUGUCUAAA AAAGACAAAAAGGUACUGUGAC -5'
P-Activity	5' -CAUUCUCCUAAGAAGCUAUUAAAAUCACA-3'
T-Activity	3' -GUAAGAGGAUUCUUCGAUAAUUUUAGUGU UGGGAAAAA -5'

Table 1, RNA oligo sequences: Listed are the sequences of the RNAs used for *in vitro* assays. Primer RNAs "P" are listed 5'-3' and contain a 5' 6-FAM modification and template RNAs "T" are listed 3'-5'. **Bolded** regions indicate 5' overhangs (T RNA) or 3' mismatches (P RNA) that remain as ssRNA in duplexes.

In vitro Primer Extension Assay:

Complexes were assembled at the following concentrations: nsp7 – 1.5 μ M, nsp8 – 1.5 μ M, nsp12 – 0.5 μ M, and dsRNA – 250 nM. Reactions were carried out in buffer conditions of 10 mM Tris-Cl, 10 mM NaCl, 2 mM MgCl₂, and 2 mM DTT. Prior to use, proteins were diluted in reaction buffer and then combined and incubated at 25°C for 15 minutes, after which dsRNA was added and incubated at 25°C for 15 minutes. Addition of NTPs at a final concentration of 40 μ M initiated primer extension. Reactions were halted by addition of two reaction volumes of denaturing loading buffer (95% (v/v) formamide, 2 mM ethylenediaminetetraacetic acid (EDTA) and 0.75 mM bromophenol blue). Samples were heated at 95°C for 15 minutes before analysis via denaturing urea-PAGE (8 M urea, 15% PAGE) run in 1X TBE (89 mM Tris-Cl, pH 8.3, 89 mM boric acid, 2 mM EDTA). Gels were imaged on a Typhoon FLA 9200 (GE Healthcare) using excitation of 470 nm and emission of 530 nm.

In vitro Exonuclease Assay:

Proteins and RNA (ss or ds) were combined at final concentrations of nsp10 - 500 nM, nsp14 - 125 nM, and RNA - 250 nM. Reaction conditions were 10 mM Tris-Cl, 10 mM NaCl, 2 mM MgCl₂, and 2 mM DTT. Proteins were diluted in reaction buffer after which proteins and RNA were combined and incubated at 30°C for 30 minutes. Reactions were halted by the addition of two reaction volumes of denaturing loading buffer (95% (v/v) formamide, 2 mM EDTA, and 0.75 mM bromophenol blue). Samples were then analyzed following an identical protocol as the “*In vitro* Primer Extension Assay”.

Electrophoretic Mobility Shift Assay (EMSA):

Reaction buffer conditions were 25 mM HEPES, pH 7.5, 50 mM NaCl, 2 mM MgCl₂, 2 mM DTT. Unless noted otherwise, protein and RNA concentrations were as follows: nsp7 – 2.0 μm, nsp8 – 3.0 μm, nsp12 – 1.0 μm, dsRNA – 1.2 μm, nsp10 – 4.5 μm, and nsp14 – 3.0 μm. Prior to use, proteins were diluted in reaction buffer. To assemble complexes, nsp7, nsp8, and nsp12 were combined and incubated at 25°C for 15 minutes, then dsRNA was added, and reactions incubated at 25°C for 15 minutes. Following this, nsp10 and nsp14 were added (if present) and reactions incubated for 15 minutes at 25°C. Following assembly, reactions were mixed with 10X native-PAGE loading buffer (10 mM Tris, 1 mM EDTA, 0.75 mM bromophenol blue, and 50% glycerol (v/v)) and loaded onto a 4.5% native-PAGE gel and run in 1X TBE buffer. The native-PAGE gel, running buffer, and gel rig were pre-chilled on ice for at least one hour before running and then ran on ice. Gels were imaged on a Typhoon FLA 9200 (GE Healthcare) using excitation of 470 nm and emission of 530 nm.

Results:

Assembly of SARS-CoV-2 Polymerase and Exonuclease Complexes:

To study complex interactions during SARS-CoV-2 proofreading, we recombinantly expressed and purified the proteins of the core-RTC (nsp7, nsp8, and nsp12) and exonuclease complex (nsp10 and nsp14) (**Fig 2A** and **S1A**). To test the activity of nsp14, we tracked the degradation of ssRNA (P-activity RNA, **Table 1**) or dsRNA (P-activity + T-activity RNA duplex, **Table 1**) +/- nsp10 (**Fig 2B** and **S1B**). Nsp14 required nsp10 to degrade both ssRNA and dsRNA, with a slight preference for ssRNA (**Fig 2B**). As described above, there are contradictory reports of the exonuclease complexes' preference for ss or dsRNA (50,51). These differing results come from several

different labs, indicating that the altered preferences could be a result of the different protein purification and assay protocols. To test the polymerase activity of the core RTC, a short dsRNA template (P-activity + T-activity, **Table 1**) was used with the core-RTC to extend the primer to the length of the template (38 nt) in the presence of NTPs (**Fig. 2C** and **S1C**). Core-RTC activity required the presence of nsp7, nsp8 and nsp12, matching prior reports of CoV RdRP activity (22,27). In addition to wildtype enzymes, we prepared catalytic knockouts of both nsp12 (Asp760Ala) and nsp14 (His268Ala) as negative controls for activity assays (**Fig. 2** and **S1**).

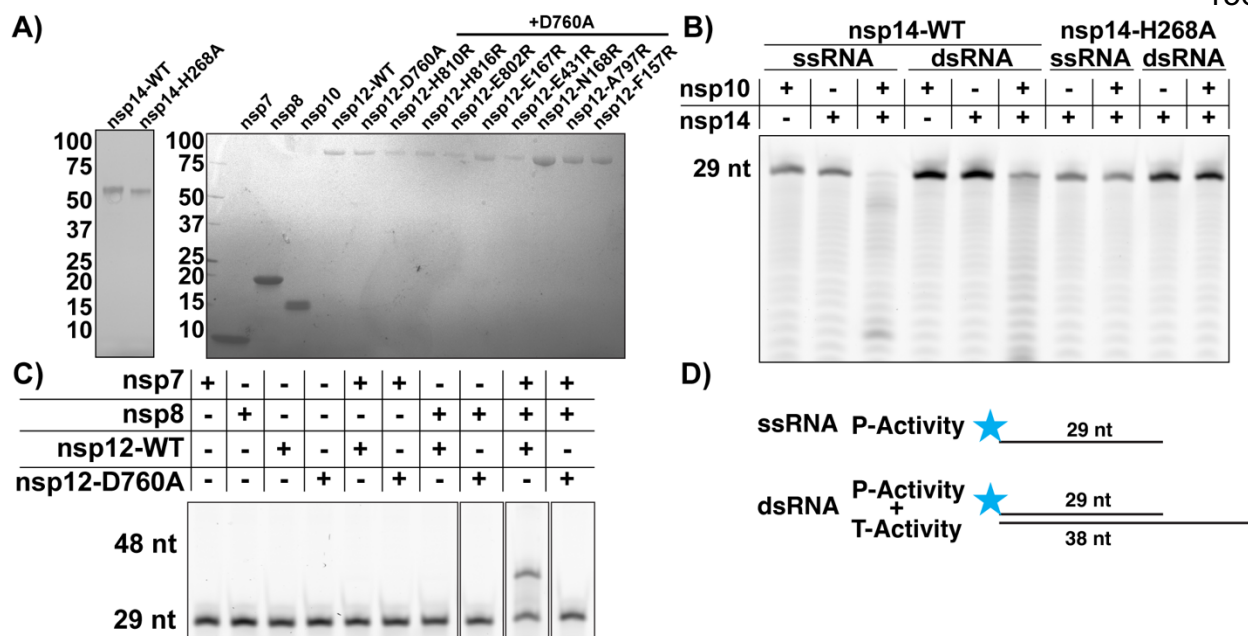


Figure 2, Assembly of active SARS-CoV-2 enzyme complexes: **A)** SDS-PAGE analysis of recombinant SARS-CoV-2 nsp7 (9,300 Da), nsp8 (21,900 Da), nsp10 (14,800 Da), nsp14 (59,800 Da), and nsp12 (110,100 Da) including nsp12 mutants used in this study. Numbers at left indicate MW (kDa) position. **B)** SARS-CoV-2 exoribonuclease complex degrades ss and dsRNA. **C)** The SARS-CoV-2 core-RTC extends an dsRNA duplex in the presence of NTPs. **D)** (ss/ds) RNAs used in **B** and **C** *in vitro* assays are diagrammed. The blue stars represent the 6-FAM fluorescent tag on the primer 5' end.

Assembly of the CoV Proofreading-RTC Requires a Backtracked Substrate:

To test exonuclease and core-RTC interactions we used an electrophoretic mobility shift assay (EMSA) that tracks the migration of fluorescently labelled RNA probes on native-PAGE gels. Using an elongation competent RNA duplex (P0 + T-backtrack, **Table 1**) we added the components of the core-RTC (nsp7, nsp8, and nsp12) and observed an upward shift in the RNA band, indicating core-RTC binding of RNA (**Fig. 3A** and **S2A**). We next added nsp14-His268Ala (+/- nsp10) and did not observe a band super shift, indicating that the exonuclease did not interact with the core-RTC (**Fig. 3A** and **S2A**). A possible explanation for the lack of nsp14 and core-RTC interactions compared to previous reports is that we used wildtype nsp7 and nsp8 rather than a nsp7-nsp8 fusion protein that may promote or force interaction (21).

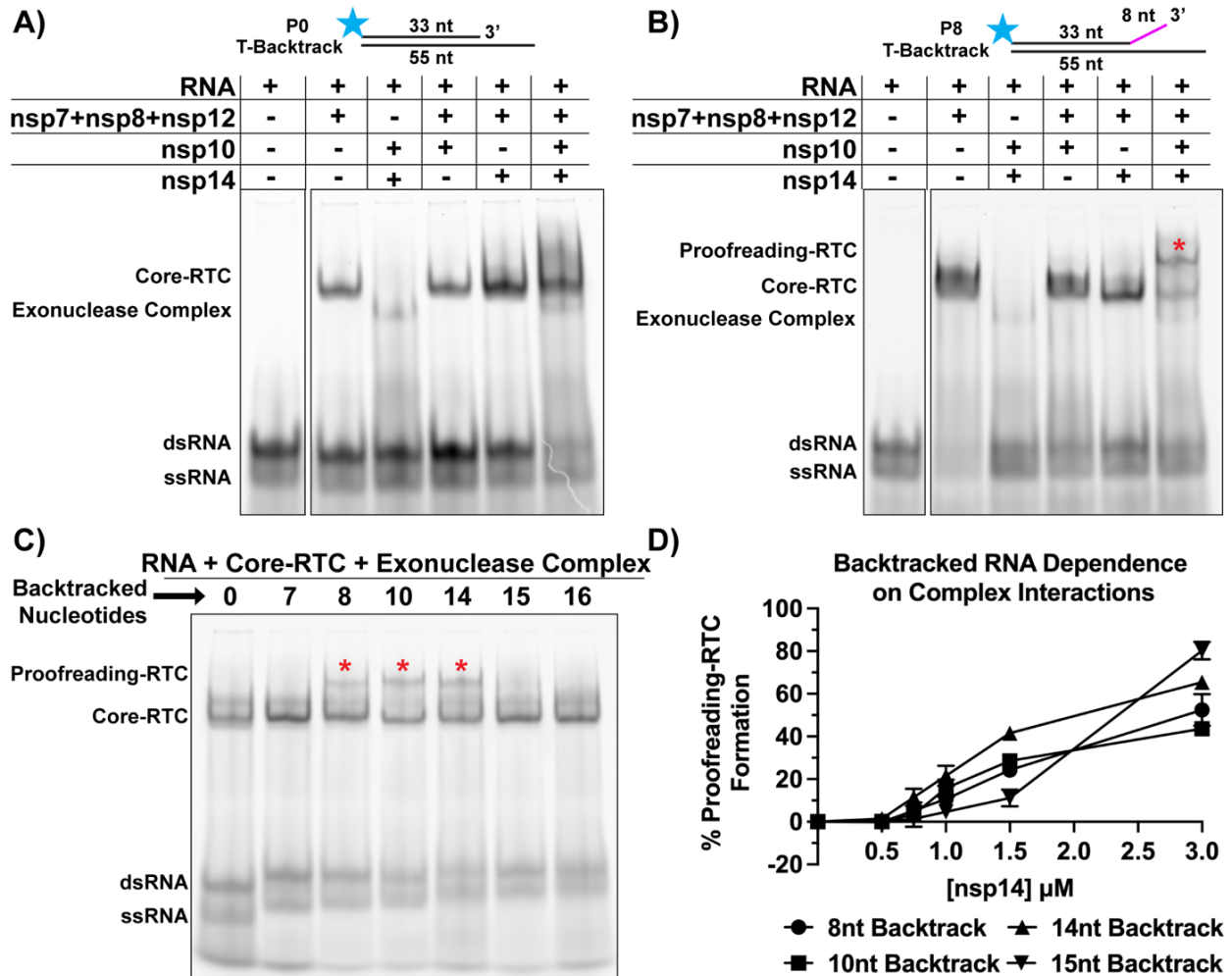


Figure 3, Backtracked RNA promotes proofreading-RTC formation: EMSAs testing core-RTC and exonuclease complex interactions on **A)** an elongation competent RNA duplex or **B)** a duplex with 8 nucleotides of backtracked RNA. **C)** Evaluation of proofreading-RTC formation at varying backtracked RNA lengths. Red asterisks in **(B)** and **(C)** indicate the super-shifted band of the proofreading-RTC. **D)** Comparison of proofreading-RTC formation at 8, 10, 14, and 15 nucleotides of backtracked RNA, tested by titrating exonuclease complex onto formed core-RTCs. For **(D)** the concentrations of the core-RTC components (nsp7, nsp8, nsp12, and duplex RNA) were halved compared to those described in the Methods and used for **(A-C)**.

In the backtracked RTC structure, the 3' end of nascent RNA is extruded through nsp12's NTP channel (**Fig. 1**) (30). We hypothesized that RTC backtracking could allow nsp14 to access 3' mismatches for proofreading that would be otherwise inaccessible in the RdRP active site. Using the backtracked structure, we modeled 8 nucleotides (nts) as the minimum amount of backtracked RNA needed to span the distance from the nsp12 polymerase active site to nsp14's exonuclease active site if nsp14 was positioned near the polymerase NTP channel (30). The nsp12 mutation Asp760Ala has been previously shown to promote backtracked RNA binding in the absence of nsp13 (30). Using an RNA duplex with 8 nts of mismatched ssRNA on the primer's 3' end (P8 + T-backtrack, **Table 1**) we assembled the core-RTC with both wildtype and mutant nsp12 (Asp760Ala) (**Fig. S3**). While wildtype nsp12 could bind the forked RNA substrate, nsp12-Asp760Ala had slightly improved RNA binding (**Fig. S3**).

Using the 8 nt backtracked RNA duplex (P8 + T-backtrack, **Table 1**) in our EMSA we observed a band super shift in the presence of the core-RTC and intact exonuclease complex (nsp10+nsp14) (**Fig. 3B** and **S2B**). These results demonstrate that the SARS-CoV-2 core-RTC and exonuclease complex interact when the RTC is in backtracked state. We refer to this isolated complex as the proofreading-RTC as we believe this to be a snapshot of CoV proofreading. To our knowledge this is the first report of an isolated coronavirus proofreading complex assembled without the use of nsp fusions. To ensure that the enzyme mutations did not produce experimental artifacts we successfully assembled the proofreading-RTC with wildtype nsp12 or nsp14 (**Fig. S3** and **S4**). To isolate the complex with wildtype nsp14, Ca^{2+} replaced Mg^{2+} in the reaction buffer to inhibit exonuclease activity (**Fig. S4**). These results support a model of CoV proofreading during

which the RTC remains bound to the RNA during mismatch excision by the exonuclease. Similar mechanisms of proofreading that involve the movement of a mis-incorporated nucleotide from the polymerase active site to a secondary exonuclease active site are used by DNA polymerases with nsp14-homologous DE(E/D)D exonucleases (33).

Backtrack Length Limitations for Complex Assembly:

To test possible backtracking constraints on proofreading-RTC assembly we designed a series of RNA primers with 4 – 18 nucleotides of 3' mismatches (**Table 1**). The core-RTC could bind each of these backtracked RNAs (**Fig. S5**). Using these model-backtrack varying-length primers, we identified a window of proofreading RTC formation of 8-14 nts of backtracked RNA (**Fig. 3C** and **S5**). We then compared the affinity of the core-RTC and exonuclease complex on four of these backtracked RNAs (P8, P10, P14, P15, **Table 1**) by titrating the exonuclease complex onto the core-RTC (**Fig. 3D** and **S6**). Although the improved binding is modest, the 14-nucleotide backtracked RNA promoted proofreading-RTC assembly the best (**Fig. 3D**). At high concentrations of the exonuclease complex, we observed proofreading-RTC assembly with 15 nts of backtrack, indicating that while complex interactions can occur with 15 backtracked nts they are disfavored. These results indicate that the interaction site of the exonuclease complex on the core-RTC might not be a rigid fit at the NTP channel. Rather, the interaction could have some plasticity, perhaps allowing nsp14 to bind the 3' end of RNA as nsp13 mediates backtracking in bursts of helicase activity (61,62).

Interaction of the Two Complexes Occurs Away from the NTP Channel:

The initial design of our 8 nt backtracked primer (P8, **Table 1**) was based on our hypothesis that the exonuclease complex binds the core-RTC at the NTP channel. But the range of backtracked RNA lengths that promote proofreading-RTC assembly suggest that interaction might not be a rigid fit at the NTP channel (**Fig 3**). To test if the exonuclease complex is binding at the NTP channel we produced eight nsp12 mutations surrounding the NTP channel (**Fig. 2A** and **4A**). We mutated highly conserved residues (Phe157, Glu167, Asn168, Glu431, Ala797, Glu802, His810, His816) to arginine to potentially disrupt protein-protein interactions at the NTP channel (**Fig. 4A**). All eight of the mutant nsp12s could bind backtracked RNA and form the proofreading-RTC (**Fig. 4B** and **S7**). Three of the nsp12 mutants (Glu167Arg, Glu431Arg, Glu802Arg) had reduced RNA binding but still formed the proofreading complex (**Fig. 4B** and **S7**). These results indicate that the interaction site between the core-RTC and the exonuclease complex is somewhere other than the NTP channel, but still positioned close enough that it could bind with 8 nucleotides of backtracked RNA. Prior publications have noted that nsp8 can interact with the proofreading complex, so the interaction may occur at a binding site with nsp8 or the nsp7+nsp8 heterodimer (59).

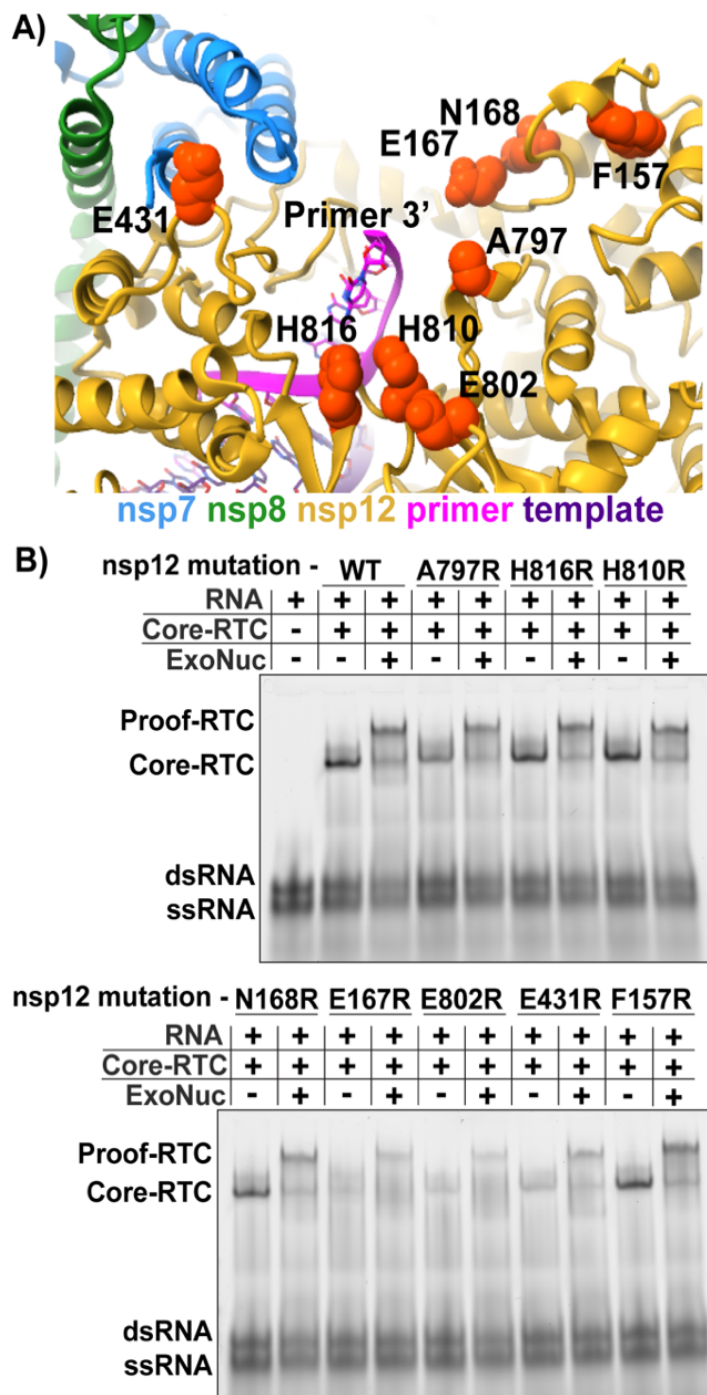


Figure 4. NTP channel mutations do not disrupt proofreading-RTC formation: A) Structure of the SARS-CoV-2 RTC in a backtracked state (PDB: 7KRP) with residues targeted for mutation highlighted as orange spheres. B) EMSA testing core-RTC and proofreading-RTC formation on a 14-nucleotide backtracked RNA duplex (P14 + T-backtracking, Table 1) with nsp12 mutations surrounding the NTP channel. Each conserved nsp12 residue was mutated to arginine. For these EMSAS nsps and duplex RNA concentrations were halved from what is described in the Methods section. Each nsp12 used also had the D760A mutation.

Discussion:

We have identified direct interactions between the CoV core-RTC and exonuclease complex and shown that proofreading-RTC assembly requires an RNA backtracked RTC. These data begin to define the mechanism of CoV proofreading. Interestingly, the assembly of the proofreading-RTC has a range of backtracked RNA length for complex assembly of 8-14 nucleotides. While we initially predicted that the exonuclease binding site would be at nsp12's NTP channel, this observation led us to hypothesize that the interaction might not be a rigid fit at the NTP channel. To our surprise, our mutational screen of potential core-RTC and exonuclease complex disrupting mutations around the NTP channel did not block proofreading-RTC assembly, revealing interaction occurs elsewhere on the RTC.

Nsp13's ability to backtrack the RTC indicates that it could be involved in CoV proofreading (30). The backtracked RNA window of interaction (8-14 nts) aligns well with nsp13's non-processive nature, as bursts of nsp13s' helicase activity are typically ~10 base pairs (61,62). While isolated nsp13 has been studied *in vitro*, whether its enzymatic parameters are altered in the context of an assembled RTC is unknown. Nsp13's backtracking bursts and the tight window of RTC-exonuclease interaction provide a mechanism to restrict nsp14's access to free 3' ends of RNA, preventing unnecessary degradation of nascent RNA. The movement of misincorporations to an exonuclease active site is a conserved mechanism of proofreading among eukaryotic and prokaryotic DNA polymerases (33). As noted above, nsp14 is in the same DE(E/D)D superfamily of exonucleases as those involved in DNA proofreading (47).

Recent work from our lab and others has shown that nsp8_T's N-terminal RNA binding extension is not required for RNA synthesis *in vitro* (22,27). Nsp8_T scaffolds the backtracking nsp13 to the RTC (**Fig. 1**) (30). We now propose that nsp8_T functions as a regulator of CoV backtracking and proofreading (**Fig. 5**). In this mechanism, nsp8_T or just its RNA binding domain, could be unengaged with the core-RTC during elongation. After a mismatch is introduced, the frayed RNA substrate would pause elongation, allowing nsp8_T and its N-terminal extension to bind the dsRNA, further stalling the RTC and allowing nsp13 to bind (**Fig. 5**). Nsp13 could then mediate a burst of backtracking to extrude the 3' mismatch out the NTP channel. The free 3' mismatch would then be bound and excised by the exonuclease complex. After excision, nsp10+nsp14 could disassociate from the RTC, allowing RNA to re-enter the active site and elongation to continue. Although nsp13 is known to backtrack the CoV RTC, nsp13's involvement in proofreading remains untested. Because our data indicates that nsp14 does not bind at the nsp12 NTP channel, the nsp7+nsp8_T heterodimer could also contribute to scaffolding the exonuclease complex to the RTC. In support of this, it has been demonstrated that nsp8 can bind nsp10+nsp14 and promote exonuclease activity (59). It should be noted that reports of this interaction do not provide clear results, and further evidence is needed to support this claim.

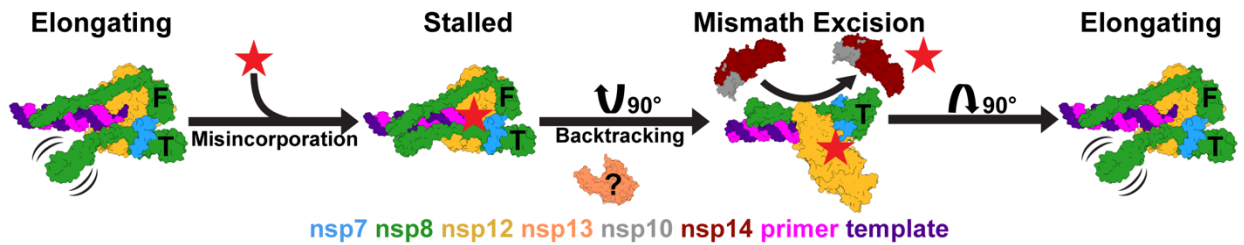


Figure 5, Hypothetical model of CoV proofreading: A misincorporation (red star) is hypothesized to initiate stalling of the CoV core-RTC. Stalling is predicted to promote nsp8_T's N-terminal extension binding to dsRNA and allow backtracking by the RTC. While nsp13 is known to mediate CoV backtracking, whether it is involved in proofreading is currently unknown. Backtracking by the CoV RTC results in the 3' end of nascent RNA to be extruded out the NTP channel, allowing the exonuclease complex to bind the RTC, and excise the mismatch. After mismatch excision, the exonuclease complex disassociates from the RTC allowing RNA to re-enter the active site and elongation to continue.

Our results indicate that the structure of a SARS-CoV-2 RTC interacting with nsp13, nsp14, and a nsp9-nsp10 fusion protein might not be representative of CoV proofreading but perhaps other steps in CoV replication (60). In the monomeric form of this RTC, the exonuclease complex primarily interacts with nsp12's NiRAN domain as a result of the nsp9-nsp10 fusion (60). This positions nsp14's exonuclease active site too far away from the NTP channel to bind 8-14 nts of backtracked RNA. Interestingly, in the dimeric form of the RTC, the exonuclease active site of one monomer is positioned *in trans* near the NTP channel of the second monomer. While this positioning is potentially more feasible to accomplish CoV proofreading, the exonuclease active site is oriented away from the NTP channel making interaction between the 3' end of RNA and nsp14's active site difficult. In addition, several of the mutants we screened that did not disrupt complex formation are at (nsp12-Glu431Arg) or near (nsp12-Glu802Arg, His810Arg, His816Arg) the nsp12-nsp14 interface of the dimeric RTC, indicating that this might not be the true binding site.

The design of effective nucleoside analogue (NA) antivirals against CoVs has proven difficult due to their proofreading ability. To be effective a NA must both be incorporated by the polymerase and evade excision by nsp14. Based on our data and hypothetical proofreading model, the latter of these (evading nsp14) could take many different forms. First, the NA could be designed to be a non-excisable substrate for nsp14, such as a NA that lacks a 2'OH. The NA could also prevent sensing by the polymerase complex, completely avoiding pausing, backtracking and proofreading altogether. One such antiviral already exists with the hyper-mutagen Molnupiravir (63). Based on the window of interaction between the exonuclease complex and RTC, antivirals could avoid

proofreading by taking effect several nucleotides (>15) after their incorporation. This window of effectiveness could explain why the NA Remdesivir, which stalls replication 3-4 nucleotides after incorporation, has increased efficacy against proofreading deficient CoVs (64). A final possibility would be drugs that block CoV backtracking by blocking RNA extruding through the NTP channel or inhibiting nsp13.

While our work provides critical insight into the mechanisms of CoV proofreading, there is still much we do not know. Our mutagenesis suggests a site other than the NTP channel as the binding site of the exonuclease complex, but the actual binding site remains unknown. Further structural and biochemical analysis, including mutagenic screens could identify this binding site. In addition, the requirement for backtracking in proofreading has not been established. Testing this using reverse genetics would be difficult as nsp13's helicase activity is essential for virus replication. It would be possible to screen nsp8 mutants that block the backtracking nsp13's ability to bind the RTC. In addition, the development of a novel *in vitro* proofreading assay would aid in parsing out the role of each protein in CoV proofreading.

References:

1. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res*, **117**, 17-37.
2. Cui, J., Li, F. and Shi, Z.-L. (2019) Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol*, **17**, 181-192.
3. de Wit, E., van Doremalen, N., Falzarano, D. and Munster, V.J. (2016) SARS and MERS: recent insights into emerging coronaviruses.
4. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.

5. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. and Fouchier, R.A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, **367**, 1814-1820.
6. de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R.A., Galiano, M., Gorbalenya, A.E., Memish, Z.A. *et al.* (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, **87**, 7790-7792.
7. Rabaan, A.A., Al-Ahmed, S.H., Sah, R., Alqumber, M.A., Haque, S., Patel, S.K., Pathak, M., Tiwari, R., Yattoo, M.I., Haq, A.U. *et al.* (2021) MERS-CoV: epidemiology, molecular dynamics, therapeutics, and future challenges. *Ann Clin Microbiol Antimicrob*, **20**, 8.
8. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265-269.
9. (2024) WHO COVID-19 dashboard.
10. Vlasova, A.N., Diaz, A., Dامتie, D., Xiu, L., Toh, T.H., Lee, J.S., Saif, L.J. and Gray, G.C. (2022) Novel Canine Coronavirus Isolated from a Hospitalized Patient With Pneumonia in East Malaysia. *Clin Infect Dis*, **74**, 446-454.
11. Silva, C.S., Mullis, L.B., Pereira, O., Jr., Saif, L.J., Vlasova, A., Zhang, X., Owens, R.J., Paulson, D., Taylor, D., Haynes, L.M. and Azevedo, M.P. (2014) Human Respiratory Coronaviruses Detected In Patients with Influenza-Like Illness in Arkansas, USA. *Viol Mycol*, **2014**, 1-8.
12. Xiu, L., Binder, R.A., Alarja, N.A., Kochek, K., Coleman, K.K., Than, S.T., Bailey, E.S., Bui, V.N., Toh, T.H., Erdman, D.D. and Gray, G.C. (2020) A RT-PCR assay for the detection of coronaviruses from four genera. *J Clin Virol*, **128**, 104391.
13. Lednicky, J.A., Tagliamonte, M.S., White, S.K., Elbadry, M.A., Alam, M.M., Stephenson, C.J., Bonny, T.S., Loeb, J.C., Telisma, T., Chavannes, S. *et al.* (2021) Independent infections of porcine deltacoronavirus among Haitian children. *Nature*, **600**, 133-137.
14. Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, **1282**, 1-23.
15. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z. *et al.* (2020) Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, **183**, 730-738 e713.
16. Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.

17. Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. and Howard, M.T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, **332**, 498-510.
18. Plant, E.P. and Dinman, J.D. (2008) The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci*, **13**, 4873-4881.
19. Malone, B., Urakova, N., Snijder, E.J. and Campbell, E.A. (2021) Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*.
20. van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog*, **4**, e1000054.
21. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
22. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.
23. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.T., Meng, X., Xia, H., Wood, J.M., Harris, L.D. *et al.* (2021) Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife*, **10**.
24. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
25. Hillen, H.S., Kobic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
26. Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C. *et al.* (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*, **182**, 417-428 e413.
27. Anderson, T.K., Hoferle, P.J., Chojnacki, K.J., Lee, K.W., Coon, J.J. and Kirchdoerfer, R.N. (2024) An alphacoronavirus polymerase structure reveals conserved replication factor functions. *Nucleic Acids Res*.
28. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T. *et al.* (2020) Structural Basis

- for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
29. Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M. *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun*, **11**, 5874.
 30. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, **118**.
 31. Wang, D., Bushnell, D.A., Huang, X., Westover, K.D., Levitt, M. and Kornberg, R.D. (2009) Structural basis of transcription: backtracked RNA polymerase II at 3.4 angstrom resolution. *Science*, **324**, 1203-1206.
 32. Abdelkareem, M., Saint-André, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J. and Weixlbaumer, A. (2019) Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell*, **75**, 298-309.e294.
 33. Buchel, G., Nayak, A.R., Herbine, K., Sarfallah, A., Sokolova, V.O., Zamudio-Ochoa, A. and Temiakov, D. (2023) Structural basis for DNA proofreading. *Nat Commun*, **14**.
 34. Mönttinen, H.A.M., Bicep, C., Williams, T.A. and Hirt, R.P. (2021) The genomes of nucleocytoplasmic large DNA viruses: viral evolution writ large. *Microb Genom*, **7**.
 35. Holmes, E.C. (2003) Error thresholds and the constraints to RNA virus evolution. *Trends Microbiol*, **11**, 543-546.
 36. Lauring, A.S. and Andino, R. (2010) Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog*, **6**, e1001005.
 37. Fitzsimmons, W.J., Woods, R.J., McCrone, J.T., Woodman, A., Arnold, J.J., Yennawar, M., Evans, R., Cameron, C.E. and Lauring, A.S. (2018) A speed-fidelity trade-off determines the mutation rate and virulence of an RNA virus. *PLoS Biol*, **16**, e2006459.
 38. Domingo, E. and Schuster, P. (2016) What Is a Quasispecies? Historical Origins and Current Scope. *Curr Top Microbiol Immunol*, **392**, 1-22.
 39. Summers, J. and Litwin, S. (2006) Examining the theory of error catastrophe. *J Virol*, **80**, 20-26.
 40. Drake, J.W. and Holland, J.J. (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci U S A*, **96**, 13910-13913.

41. Crotty, S., Cameron, C.E. and Andino, R. (2001) RNA virus error catastrophe: direct molecular test by using ribavirin. *Proc Natl Acad Sci U S A*, **98**, 6895-6900.
42. Perales, C. and Domingo, E. (2016) Antiviral Strategies Based on Lethal Mutagenesis and Error Threshold. *Curr Top Microbiol Immunol*, **392**, 323-339.
43. Morais, P., Trovão, N.S., Abecasis, A.B. and Parreira, R. (2022) Readdressing the genetic diversity and taxonomy of the Mesoniviridae family, as well as its relationships with other nidoviruses and putative mesonivirus-like viral sequences. *Virus Res*, **313**, 198727.
44. Saberi, A., Gulyaeva, A.A., Brubacher, J.L., Newmark, P.A. and Gorbalenya, A.E. (2018) A planarian nidovirus expands the limits of RNA genome size. *PLoS Pathog*, **14**, e1007314.
45. Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B. and 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.
46. Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D. and Baric, R.S. (2011) Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol*, **8**, 270-279.
47. Zuo, Y. and Deutscher, M.P. (2001) Exoribonuclease superfamilies: structural analysis and phylogenetic distribution. *Nucleic Acids Res*, **29**, 1017-1026.
48. Moser, M.J., Holley, W.R., Chatterjee, A. and Mian, I.S. (1997) The proofreading domain of Escherichia coli DNA polymerase I and other DNA and/or RNA exonuclease domains. *Nucleic Acids Res*, **25**, 5110-5118.
49. Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. and 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.
50. Chinthapatla, R., Sotoudegan, M., Srivastava, P., Anderson, T.K., Moustafa, I.M., Passow, K.T., Kennelly, S.A., Moorthy, R., Dulin, D., Feng, J.Y. *et al.* (2023) Interfering with nucleotide excision by the coronavirus 3'-to-5' exoribonuclease. *Nucleic Acids Res*, **51**, 315-336.
51. Baddock, H.T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L.P., Cruz-Migoni, A., Fan, H., Keown, J.R., Walker, A.P. *et al.* (2022) Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Res*, **50**, 1484-1500.

52. Ogando, N.S., Zevenhoven-Dobbe, J.C., van der Meer, Y., Bredenbeek, P.J., Posthuma, C.C. and Snijder, E.J. (2020) The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J Virol*.
53. Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L. and Denison, M.R. (2007) High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol*, **81**, 12135-12144.
54. Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B. *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog*, **6**, e1000896.
55. Ferron, F., Subissi, L., Silveira De Moraes, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonrhein, C., Bricogne, G., Canard, B. and 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.
56. Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M. and Denison, M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog*, **9**, e1003565.
57. Smith, E.C., Case, J.B., Blanc, H., Isakov, O., Shomron, N., Vignuzzi, M. and Denison, M.R. (2015) Mutations in coronavirus nonstructural protein 10 decrease virus replication fidelity. *J Virol*, **89**, 6418-6426.
58. von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R. and Haas, J. (2007) Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFome. *PLoS One*, **2**, e459.
59. Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B. and Yang, Y. (2021) Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science*, **373**, 1142-1146.
60. Yan, L., Yang, Y., Li, M., Zhang, Y., Zheng, L., Ge, J., Huang, Y.C., Liu, Z., Wang, T., Gao, S. *et al.* (2021) Coupling of N7-methyltransferase and 3'-5' exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell*, **184**, 3474-3485.e3411.
61. Lehmann, K.C., Snijder, E.J., Posthuma, C.C. and Gorbalenya, A.E. (2015) What we know but do not understand about nidovirus helicases. *Virus Res*, **202**, 12-32.
62. Adedeji, A.O., Marchand, B., Te Velhuis, A.J., Snijder, E.J., Weiss, S., Eoff, R.L., Singh, K. and Sarafianos, S.G. (2012) Mechanism of nucleic acid unwinding by SARS-CoV helicase. *PLoS One*, **7**, e36521.

63. Sheahan, T., Sims, A., Zhou, S., Graham, R., Puijssers, A., Agostini, M.L., Leist, S., Schäfer, A., Dinnon, K.H.r., Stevens, L. *et al.* (2020) An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Science Translational Medicine*, **12**.
64. Agostini, M.L., Andres, E.L., Sims, A.C., Graham, R.L., Sheahan, T.P., Lu, X., Smith, E.C., Case, J.B., Feng, J.Y., Jordan, R. *et al.* (2018) Coronavirus Susceptibility to the Antiviral Remdesivir (GS-5734) Is Mediated by the Viral Polymerase and the Proofreading Exoribonuclease. *mBio*, **9**, e00221-00218.

Supplemental Figures:

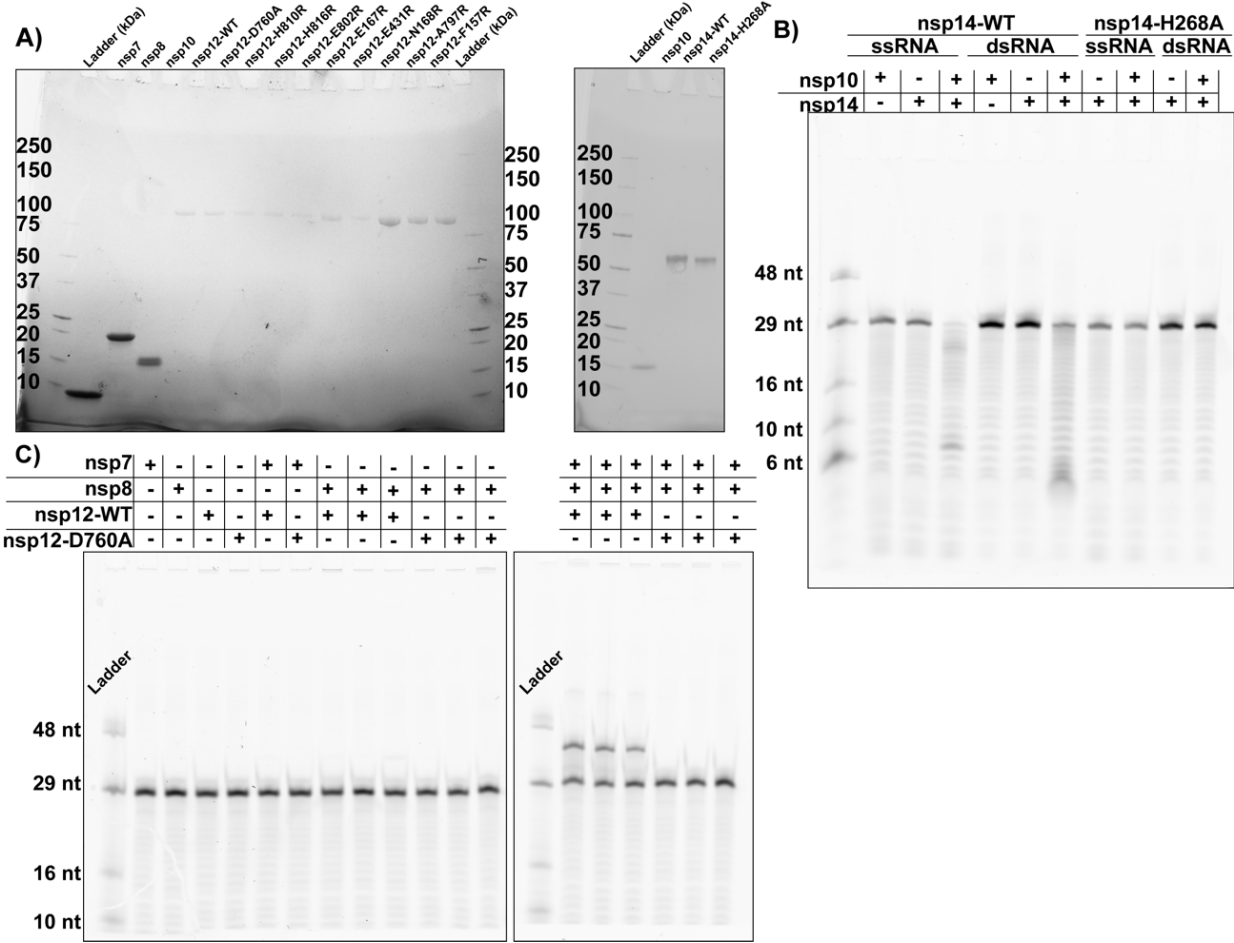


Figure S1. Full gel images for SDS-PAGE and activity assays: **A)** SDS-PAGE analysis of recombinant nsp7 (9,300 Da), nsp8 (21,900 Da), nsp10 (14,800 Da), wildtype and mutant nsp14 (59,800 Da), and wildtype and mutant nsp12 (110,100 Da). Numbers indicate MW. **B)** *In vitro* exonuclease assay tracking the degradation of ss or dsRNA in the presence of nsp10 and/or nsp14 (wildtype and H268A mutant). **C)** *In vitro* primer-extension assay testing the extension of a 29-nucleotide primer in presence of the SARS-CoV-2 core-RTC nsps. These uncropped gels were used to derive the images in Fig 2A, B, and C.

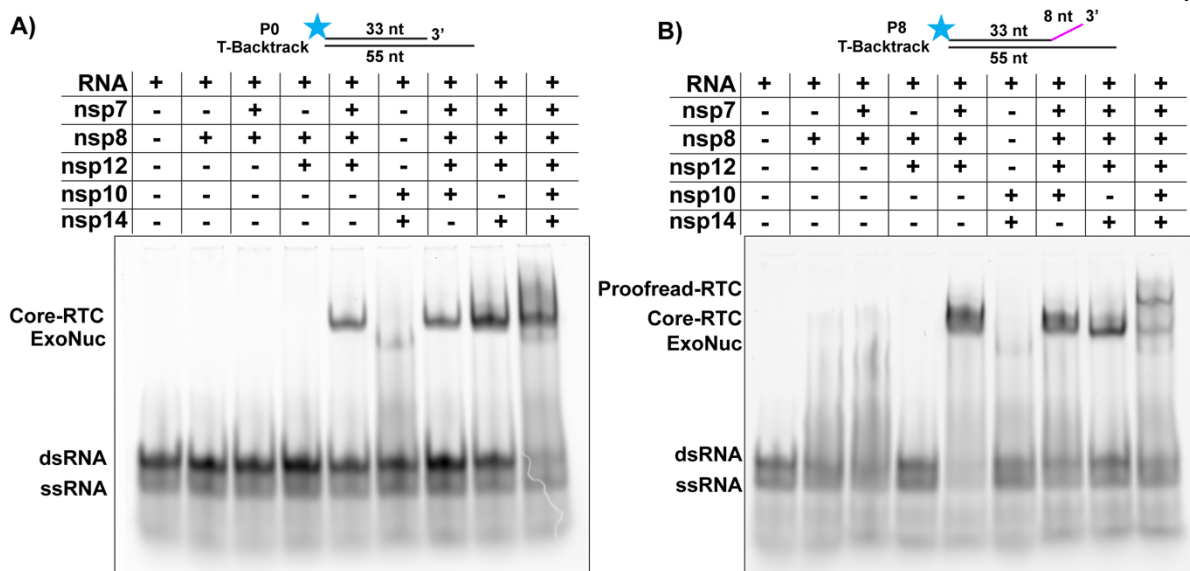


Figure S2, Full gel images for piecewise RTC EMSAs: Uncropped gels testing the RNA binding of the core-RTC, exonuclease complex, and proofreading RTC on a fully annealed RNA duplex (**A**) or backtracked RNA duplex (**B**). Nsp12-D760A was used for both EMSAs.

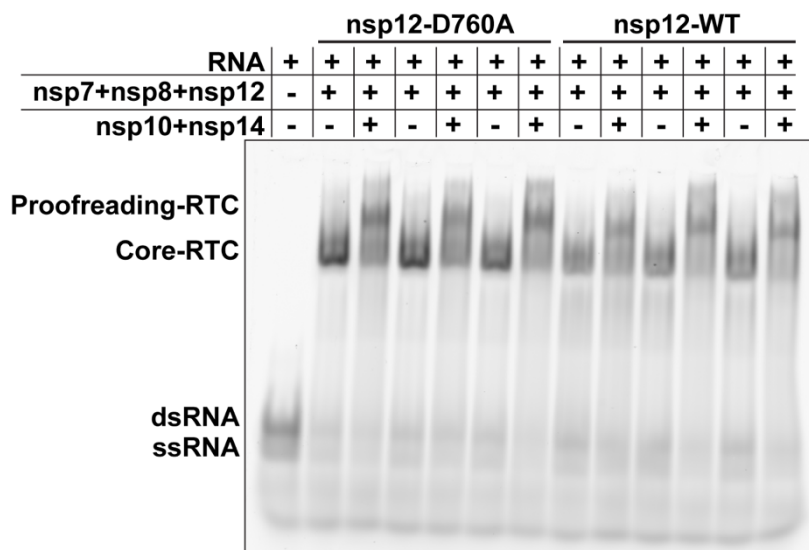


Figure S3, Mismatched RNA binding of wildtype and D760A-nsp12: EMSA comparing the ability of core-RTCs with nsp12-wildtype or nsp12-D760A to bind an RNA duplex with 8 nucleotides of mismatched RNA and form the proofreading-RTC. Reactions for each core-RTC (wildtype vs. D760A) were done in triplicate. The darker intensity bands for mutant RTCs indicate that nsp12-D760A containing RTCs bind the backtracked RNA substrate with higher affinity.

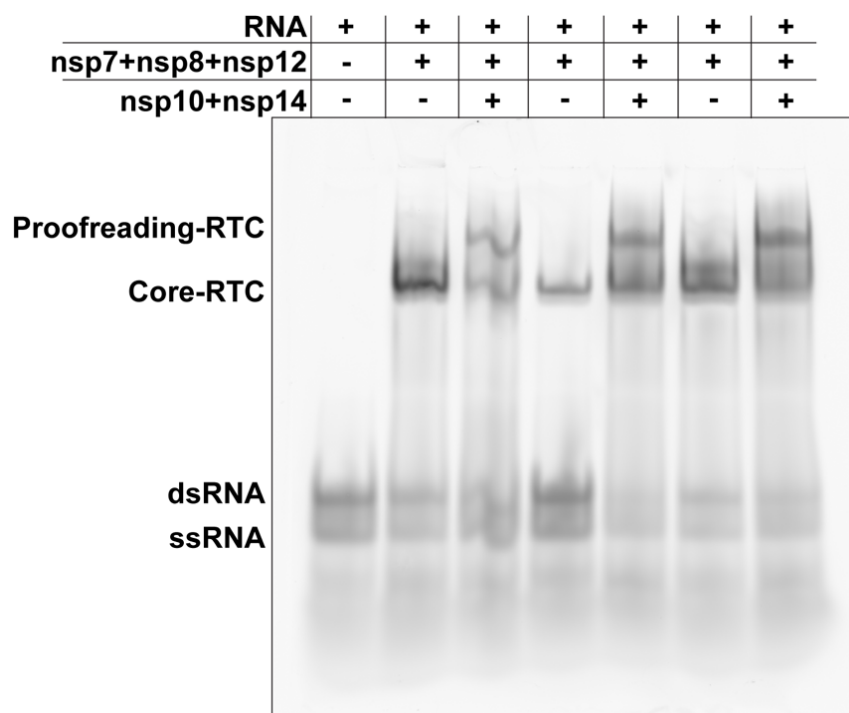


Figure S4, Proofreading-RTC assembly with wildtype nsp14: EMSA testing the formation of the SARS-CoV-2 proofreading RTC with wildtype nsp14 on an RNA duplex with 8 nucleotides of primer 3' end mismatches (P8 + T-backtrack, **Table 1**). To prevent degradation of the RNA substrate by wildtype nsp14 Mg^{2+} in assay reaction buffer was replaced with Ca^{2+} . Reactions were carried out in triplicate using nsp12-D760A.

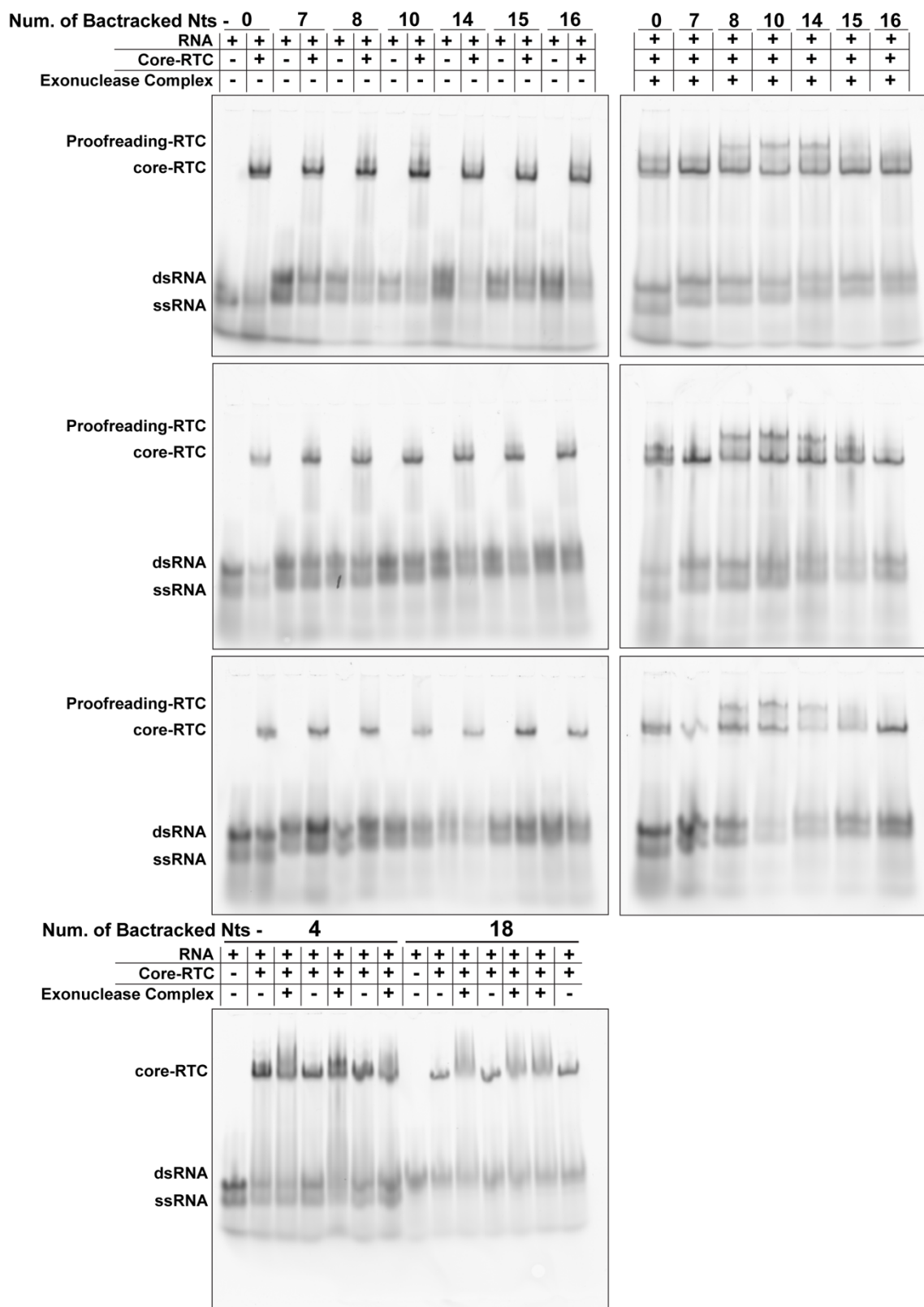


Figure S5. Full gel images for backtrack length EMSAs: EMSAs testing proofreading-RTC formation with varying lengths of primer 3' mismatches were run in triplicate.

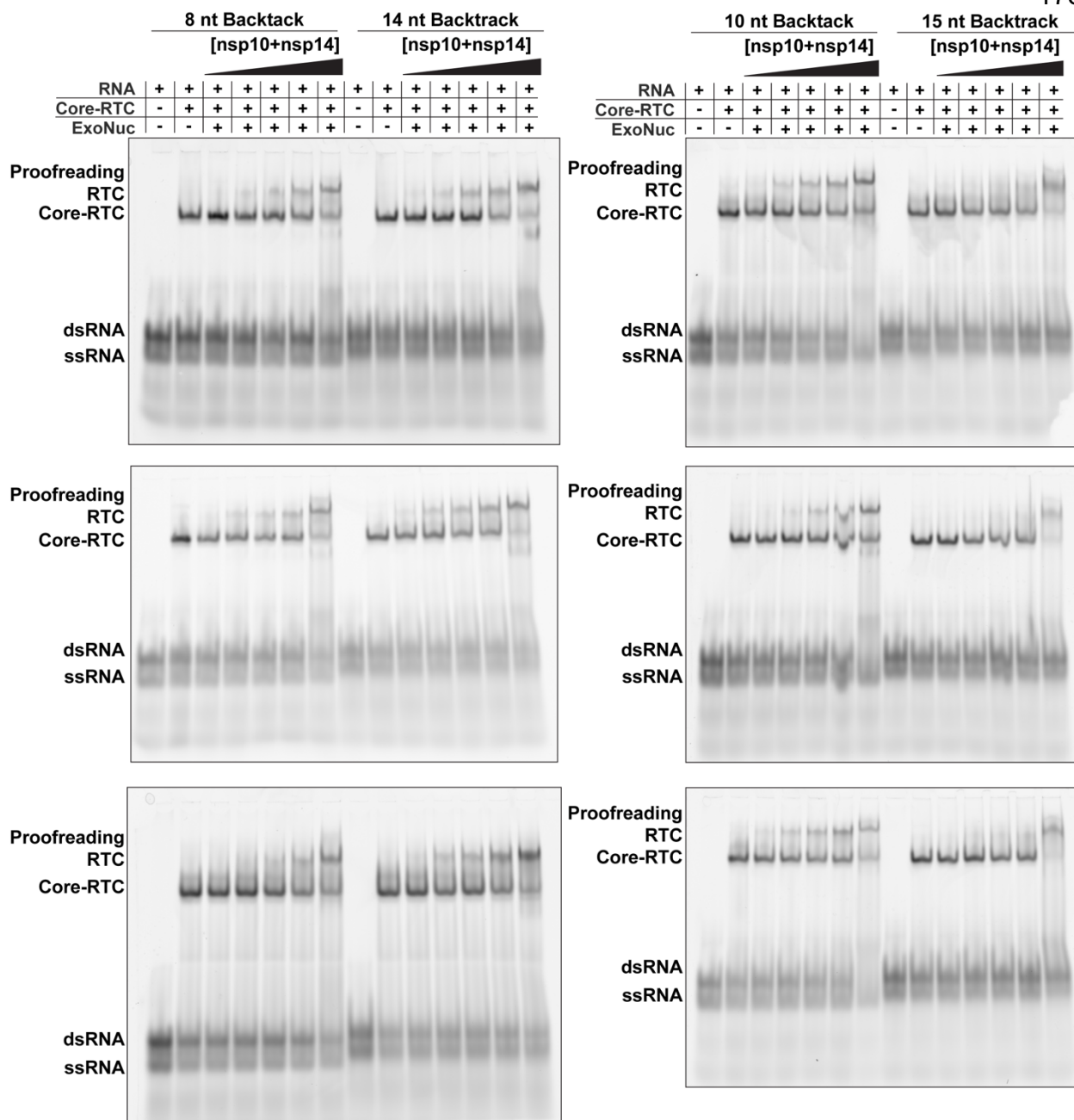


Figure S6. Exonuclease complex titration EMSAs: To determine an approximate binding comparison between the SARS-CoV-2 core-RTC and exonuclease complex with different lengths of backtracked RNA we performed titrations of the exonuclease complex onto formed core-RTCs with 8, 10, 14, or 15 nucleotides of backtracked RNA. For these experiments the core-RTC components were kept at fixed concentrations of nsp7 – 1.0 μM , nsp8 – 1.5 μM , nsp12 – 0.5 μM , and RNA duplex – 0.6 μM . Concentrations of nsp14 used to develop the curve were 0.5, 0.75, 1.0, 1.5, and 3.0 μM with nsp10 always at 1.5X nsp14's concentration. Quantitated data is presented in Figure 3D.

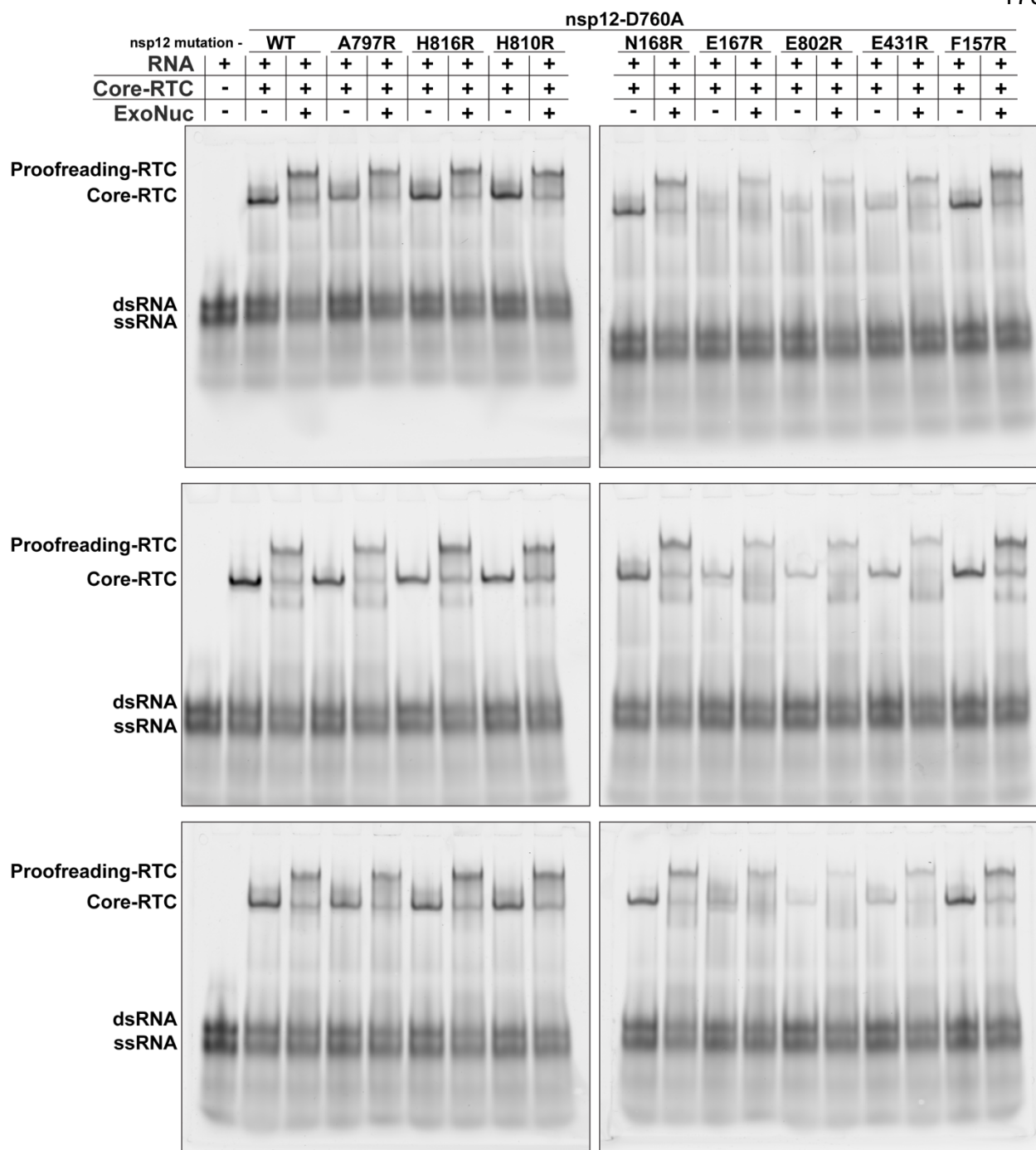


Figure S7, Full gel images for mutant nsp12: EMSAs testing proofreading RTC formation with NTP channel mutations were run in triplicate. For these mutant nsp12 EMSAs, nsps and duplex RNA concentrations were halved from what is described in the Methods section.

δ	HKU19	-LGYSSNQNNSYLNRVKGSS-DARLEPCTSDNRPDVVVRAFNINNN--ATAGIFKSTKNN	56
	HKU11	-----NSPYLNRVTGSS-GARLEPQQPGVTPDAVKRAFHVHNN--TTSIGIFLSTKTN	49
α	PDCV	-----NSAYLNRVTGSS-DARLEPLQPGTQPDVAVKRAFHVHND--TTSIGIFLSTKSN	49
	FCoV_65F	----GTTVDQSYLNRVGRSS-AARLEPCN-GTDPDHVSRAPDIYNK--DVACIGKFLKTN	52
	HCoV_229E	-----SFSSYLNRVGRSS-AARLEPCN-GTDIDYCVRAFVDVYNK--DASFIGKNLKS	50
	HCoV_NL63	-----SVDISYLNRARGSS-AARLEPCN-GTDIDKCVRAFDIYNK--NVSFLGKCLKMN	50
	HKU8	-----SLDNNYLNRVGRSS-AARLEPCN-GTEPEHVIRAFDIYNK--DVACIGKFKVKN	50
γ	PEDV	-----STDMAYLNRVGRSS-AARLEXCN-GTDTQHVYRAFDIYNK--DVACLGKFLKVN	50
	BWCoV	-----SVPQNYLNRVGRGLS-EARLRPCASGLLPDVVKRAFVDLYNS--NTAGMYASLKH	51
	IBV	SVAVASGFDKNYLNRVGRSS-EARLIPLANGCDPDVVKRAFVDVKN--ESAGMFQNLKRN	57
	HCoV_OC43	-----SKDTNFLNRVGRGASVDARLVPCASGLSTDVQLRAFDIYNA--SVAGIGLHLKVN	52
	HCoV_HKU1	-----SKDLNFLNRVGRGTSVNARLVPCASGLSTDVQLRAFDCNT--NRAGIGLYYKVN	52
β	MHV	-----SKDTNFLNRVGRGTSVNARLVPCASGLDQVQLRAFDCNA--NRAGIGLYYKVN	52
	MERS	-----SKDSNFLNRVGRGTSVNARIEPCSSGLSTDVVFRAFDCNYKAKVAGIGKYYKTN	54
	HKU4	-----SKDTNFLNRVGRGTSVNARIEPCSSGLTDDVYRAFDCNFKARVAGIGKYYKTN	54
	HKU5	-----SKDSNFLNRVGRGTSVNARIEPCASGLTDDVVFRAFDCNYKAKVAGIGKYYKTN	54
	HKU9	-----AKDECFLNRVGRGTSVAVARLVPLGSGVQPDIVLRAFDCNT--KVAGFLHLKNN	52
	SARS-CoV-2	-----SADAQSFNLNRVCGVS-AARLTPCGTGTSTDVYRAFDIYND--KVAGFAKFLKTN	52
	SARS-CoV	-----SADASTFLNRVCGVS-AARLTPCGTGTSTDVYRAFDIYNE--KVAGFAKFLKTN	52
		:***. * **:	: ***.: * :

δ	HKU19	CTRFKSTRPGSIINKPVRTIETFFVTQKCTENVFRAEEQCYNMLPKSIVSTDDKFSCVAY	116
	HKU11	CSRFRTTKQNLPLP-NKGSVELYFVSKQCSQVFEIEETCYNNMFDDSLKSTPEKFGVLAR	108
α	PDCV	CARFKTTRSALPLP-NKGEVELYFVTQCAAKVFEIEECCYNALSTELYTTDDTFGVLA	108
	FCoV_65F	CSRFRNLD-----KRDAFYVVKRCKTSVMDHEQVCYNDLKD-----SGAVAE	94
	HCoV_229E	CVRFKNVD-----KDDAFYIVKRCIKSVMDHEQSMYNLLKG-----CNAVAK	92
	HCoV_NL63	CVRFKNAD-----LKDGYFVVKRCKTSVMEHEQSMYNLLNF-----SGALAE	92
	HKU8	CVRFKNAD-----KHDAFYVVKRCKTSVMEHEQS IYDALKD-----CGAVSP	92
γ	PEDV	CVRLKNLD-----KHDAFYVVKRCKTSAMEHEQS IYSRLEK-----CGAVAE	92
	BWCoV	CARFQELDENDD-----EIDSFVVKQTTPHNFHEHEKCYLDLKA-----DCVAV	96
	IBV	CARFQEVDRDTE--GNLEYCDSYFVVKQTTSPNYEHEKACYEDLKS-----EVTAD	106
	HCoV_OC43	CCRFQRVDENGD-----KLDQFVVKRTDLTIYNREMKCYERVKD-----CKFVAE	98
	HCoV_HKU1	CCRFQRIDDDGN-----KLDKFFVVKRTNLEVNKEKTYELTKS-----CGVVAE	98
β	MHV	CCRFQRVDDEGN-----KLDKFFVVKRTNLEVNKEKECYELTKE-----CGVVAE	98
	MERS	TCRFVELDDQGH-----HLDSYFVVKRHTMENYELEKHCYDILLRD-----CDAVAP	100
	HKU4	TCRFVQVDDEGH-----KLDSYFIVKRHTMSNYELEKRCYDILLKD-----CDAVAI	100
	HKU5	TCRFVEVDDEGH-----RLDSYFVVKRHTMENYELEKRCYDILVKD-----CDAVAV	100
	HKU9	CCRYQELDADGN-----QLDSYFVVKRHTESNYLLEQRCYELKLD-----CDVVAR	98
	SARS-CoV-2	CCRFQEKDEDDN-----LIDSYFVVKRHTFSNYQHEETIYNLLKD-----CPAVAK	98
	SARS-CoV	CCRFQEKDEEGN-----LLDSYFVVKRHTMSNYQHEETIYNLVKD-----CPAVAV	98
		* : : : * :	* * :

NIRAN
Motif A_N

δ	HKU19	HDFFKFDG----VNNVVRRLTKYTLTLLDLVYALRHLSTS-QEIIQEILTTMCGTSEE---	168
	HKU11	TEFFKFDK----IPNVNRQFLTKYTLTLLDLAYALRHLSTS-RDVIKEILITICGTTEE---	160
α	PDCV	TEFFKFDK----IPNVNRQYLTKYTLTLLDLAYALRHLSTS-KDVIQEILITMCGTPED---	160
	FCoV_65F	HDFFLYKEGRCEFGNVARKDLTKYTMDLCLYAIRNFDEKNCEVLKEILVTLGACNES---	151
	HCoV_229E	HDFFTWHEGRTIYGNVSRQDLTKYTMDLCLFALRNFDEKDCGVFKEILVLTGCCNTD---	149
	HCoV_NL63	HDFFTWKDGRVIYGNVSRHNLTKYTMDLVYAMRNFDENQCDVLKEVLVLTGCCDNS---	149
	HKU8	HDFFWKDGSRVYGN IARHDLTKYTMDLVYHALRNFDEKNCEVLKEILVLSGACDSS---	149
γ	PEDV	HDFFTWKDGRAIYGNVCRKDLTEYTMDLCLYALRNFDENNCDVLKSLIKVGCCEES---	149
	BWCoV	HDFFRFE-G----MYSICRQLTKYTMDLCLYAFRHFDPNDCDVLKEILVVKGCCEWD---	149
	IBV	HDFVFVNKN---IYNISRQLTKYTMDLCLYALRHFDPKDCVLELKVYGCIEIDYHPK	163
	HCoV_OC43	HDFFTFDVSGSRVPHIVRKDLTKYTMDLCLYALRHFDRNDCMLLCLILSYAGCEQS---	155
	HCoV_HKU1	HDFFTFDIDGSRVPHIVRRNLSKYTMDLCLYALRHFDRNDCSILCEILCEYADCKES---	155
β	MHV	HEFFTFDVEGSRVPHIVRKDLSKFTMDLCLYALRHFDRNDCSTLKEILLTYAECEES---	155
	MERS	HDFFIQVVDKVTTPHIVRQLTEYTMDLVYALRHFQDN-SEVLKAILVKYGCCEVDT---	156
	HKU4	HDFFIQVVDKVTTPHIVRQSLTEYTMDLVYALRHFQDNQNCVLEKSLVKYGCCEQS---	157
	HKU5	HDFFIQVVDKVTTPHIVRQSLTEYTMDLVYALRHFQDNQNCVLEKSLVKYGCCEQS---	157
	HKU9	HDFFKFNIEGVMTPHVSRELRKTYTMADLVYSLRHFDRNDCSTLKEILVLRGCCTVD---	155
	SARS-CoV-2	HDFFKFRIDGDMVPHISRQLTKYTMDLVYALRHFDEGNCDTLKEILVYTNCCDDD---	155
	SARS-CoV	HDFFKFRVDGDMVPHISRQLTKYTMDLVYALRHFDEGNCDTLKEILVYTNCCDDD---	155
		: * * : : * . * : : * : * : * : * : * :	: * * :

NIRAN
Motif B_N

δ	HKU19	WF--VDGWYDFIENPTFYDEFHKLGLSINNVCVVMANKFADTCKTVGLVGLTADNQDLGG	226											
	HKU11	WF--GDSWFDPIENPTFYREFHKLGLSVLNRCVLNANAFKACSELGIVGLTTPDNQDLG	218											
	PDCV	WF--GENWFDPIENPSFYKEFHKLGLDILNRCVLNANKFASACIDAGLVGLTTPDNQDLG	218											
	FCoV_65F	-FFENKDWDFPVENEAIHEVYARLGP IVANAMLKCVAFCDAIVEKGYIGIITLDNQDLNG	210											
	HCoV_229E	-YFEMKNWDFPIENEDIHRVYAALGKVVANAMLKCVAFCDAMVKGVVGLTLDNQDLNG	208											
	HCoV_NL63	-YFDSKGWYDFVENEDIHRVYASLGKIVARAMLKCVALCDAMVAKGVVGLTLDNQDLNG	208											
	HKU8	-YFDNKNWYDFVENEDIHRVYAKLGCVVANAMLKCVAFCDAMVAKGVVGLTLDNQDLNG	208											
	PEDV	-YFNNKVVWDFPVENEIHRVYALLGPIVSRAMLKCVKFC DAMVEQGI VGVVTLTNQDLNG	208											
	BWCoV	-YFDQPNWYDFVENPDWFSLSIRLGP IFQRALIKVAEFCDLMVEKGYIGVVTLDNQDLNG	208											
	IBV	WFEENKDWYDFIENPKYIAMLAKMGP IVRALLNAIEFGNLMVEKGYVGVITLDNQDLNG	223											
	HCoV_OC43	-YFTKKDWYDFVENPDI INVYKLLGPIFNRLVSAATEFADKLVEVGLVGLTLDNQDLNG	214											
	α	HCoV_HKU1	-YFSKKDWYDFVENPDI INIYKLLGPIFNRLALLNTVIFADTLVEVGLVGLTLDNQDLYG	214										
MHV		-YFQKKDWYDFVENPDI INVYKLLGPIFNRLALLNTAKFADALVEAGLVGLTLDNQDLYG	214											
MERS		-YFENKLDWDFVENPSVIGVYHKLGERVRQA ILNTVKFCDHMKAGLVGLTLDNQDLNG	215											
HKU4		-YFDNKLWDFVENPSVIGVYHKLGERIRQAMLNTVKMCDHMKAGLVGLTLDNQDLNG	216											
HKU5		-YFDNKLWDFVENPNVISVYHKLGERIRQAVLNTVKFCDQMVKAGLVGLTLDNQDLNG	216											
HKU9		-YFDRKDWYDFVENPDI IRVYHKLGTVRKAVLSAVKADAMVEQGLIGVTLTNQDLNG	214											
SARS-CoV-2		-YFNKKDWYDFV EN PDILRVYANLGERVRQALLKTVQFC DAMRNAGIVGLTLDNQDLNG	214	SARS-CoV-2 F157R, E167R, N168R										
SARS-CoV		-YFNKKDWYDFVENPDILRVYANLGERVRQSLKTVQFC DAMRDAGIVGLTLDNQDLNG	214											
		:	*** : **	:	.. : .	:	* : * : * : * : * : * : *							
β		HKU19	QIYDFGDFVVTQPGNGCIEMDAYLSYIMPSMSMTHMLKCECLDD----NGSYKDYSIYQY	282										
		HKU11	QIYDFGDFIITQPGNGCVLDSSYYSYLMPI MSMTMHLKCECYDN----DNEIDYDGFQY	274										
		PDCV	QIYDFGDFIITQPGNGCVLDASYSYLMPI MSMTMHLKCECMDS----DGNPLEYDGFQY	274										
	FCoV_65F	NFYDFGDFVKTAPGFGACVTSYYSYMMPLMGMTSCLESENFKVSDIYGSYKQYDILLAY	270											
	HCoV_229E	NFYDFGDFVLCPPGMGIPYCTSYAYMMPVMGMTNCLASECFMKSDIFGQDFKTFDLLKY	268											
	HCoV_NL63	NFYDFGDFVLSLPMNGVPCCTSYYSYMMPI MGLTNCLASECFVKSDIFGSDFKTFDLLKY	268											
	HKU8	NFYDFGDFDTIGIPGVPLATSYSYLMPI MGMTNCLARECFVKSEIFGSDFKTYDLELY	268											
	PEDV	DFYDFGDFTCISIKMGIPICTSYYSYMMPI MGMTNCLASECFVKSDIFGDFKSYDLELY	268											
	BWCoV	NFYDFGDFKVKVLPGGVVPVTTSYYSYMMPLTACDALASERFFEFKA-TSGYKQYDLTKY	267											
	IBV	KFYDFGDFQKTAPGAGVPVFDYYSYMMPI IAMTDALAPERVFYDYV-HKGYSYDLLKY	282											
	HCoV_OC43	KWYDFGDYVIAAPGCGVAIADSYYSYIMPMLTMCHALDCELYV----NNAYRLF DLVQY	269											
	HCoV_HKU1	QWYDFGDFIQKTAPGFGVAVADSYSYMMPLTMCHVLDCELFV----NDSYRQDFDLVQY	269											
δ	MHV	QWYDFGDFVKTVPGGVAVADSYSYMMPLTMCHALDSELFV----NGTYREFDLVQY	269											
	MERS	KWYDFGDFVITQPGSGVAIVDSYYSYLMPI VLSMTDCLAAETHRDCDF-NKPLIEWPLLEY	274											
	HKU4	KWYDFGDFVITQPGAGVAIVDSYYSYLMPI VLSMTNCLAAETHRDCDF-NKPLIEWPLLEY	275											
	HKU5	KWYDFGDFVITQPGAGVAIVDSYYSYLMPI VLSMTNCLAAETHRDCDF-NKPLIEWPLLEY	275											
	HKU9	QWYDFGDFIEGPAGGVAVMDTYYSYSLAMP IYTMNTILAAECHVSGDL-CNLKRVLDIFKY	273											
	SARS-CoV-2	NWYDFGDFIQKTTPGSGVPVVDYSYLLMPL ILLTRALTAESHVDTDL-TKPIYKDWLLKY	273											
	SARS-CoV	NWYDFGDFVQVAPGCGVPIVDSYYSLLMPL ILLTRALAAESHMDADL-AKPLIKDWLLKY	273											
		.	*****:	.	*	:	**	:	**	:	*	*	:	*
	α	HKU19	DFTDYKMELFNKYFRHWSQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	342										
		HKU11	DFTDYKMELFNKYFRHWSQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	334										
		PDCV	DFTDYKMELFNKYFRHWSQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	334										
		FCoV_65F	DFTDHKEKLFNKYFKHWRDQTYHPNCSDCTSDDCI IHCANFNILFAMCLPNTCFGNLCSQA	330										
HCoV_229E		DFTEHKEVLFNKYFKYWGQDYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	328											
HCoV_NL63		DFTEHKEVLFNKYFKHWSFDYHPNCSDCYDDMCV IHCANFNILFAMCLPNTCFGNLCSQA	328											
HKU8		DFTEHKEVLFNKYFKHWRDQTYHPNCSDCYDDMCV IHCANFNILFAMCLPNTCFGNLCSQA	328											
PEDV		DFTEHKEVLFNKYFKHWRDQTYHPNCSDCYDDMCV IHCANFNILFAMCLPNTCFGNLCSQA	328											
BWCoV		DFTEEKQLFMKYNFYKWRDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	327											
IBV		DYTEEKQDLFQKYFKYWDQTYHPNCSDCYDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	342											
HCoV_OC43		DFTDYKLELFNKYFKHWSMYPHPNTVDCQDDRCI IHCANFNILFAMCLPNTCFGNLCSQA	329											
HCoV_HKU1		DFTDYKLELFNKYFKYWGMYHPNTVDCQDDRCI IHCANFNILFAMCLPNTCFGNLCSQA	329											
β	MHV	DFTDYKLELFNKYFKHWSMYPHPNTVDCQDDRCI IHCANFNILFAMCLPNTCFGNLCSQA	329											
	MERS	DFTDYKQVLFNKYFKYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	334											
	HKU4	DYTDYKIGLFNKYFKYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	335											
	HKU5	DYTDYKIGLFNKYFKYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	335											
	HKU9	YTYQFKYSLFSNFYKQYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	333											
	SARS-CoV-2	DFTEERLKLFDYFKYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	333											
	SARS-CoV	DFTEERLKLFDYFKYWDQTYHPNCVDCVDDRCIVHCANFNILFAMCLPNTCFGNLCSQA	333											
		:	** :	** .**	*	:	** :	*	:	* . :	** : ** : ** : ** : ** : ** : ** : *			

NIRAN Motif C_N

SARS-CoV-2 F157R, E167R, N168R

δ	HKU19	TVDGHPIVQTVGLHSELGIVMNDVNNHMSNINMPTLLRLVGDPTTMCVADACLDLRT	402	
	HKU11	TVDGHKI IQTVGVHLKELGIVLNQDVNTHMSNINLNTLLRLVGDPTTIASVSDKCLDFRT	394	
	PDCV	TVDGHLVVQTVGVHLKELGIVLNQDVNTHMANINLNTLLRLVGDPTTIASVSDKCVDLRT	394	
	α	FCoV_65F	HIDGVPVVVTAGYHFKQLGIVVNLVDKLDTMKLTMTDLLRFVTDPTLLVASSPALLDQRT	390
		HCoV_229E	FIDGVPVATAGYHFKQLGLVWVKDVNTHSTRLTITELLQFVTDPTLIVASSPALVDKRT	388
	α	HCoV_NL63	FIDGVPLVTTAGYHFKQLGLVWVKDVNTHSVRLTITELLQFVTDPSLIASSPALVDQRT	388
		HKU8	FIDGVPVVTAGYHFKQLGLVWVKDLNTHSTRLTINELLRFVTDPALVASSPALFDQRT	388
	γ	PEDV	WIDGVPLVTTAGYHFKQLGIVVNNLNLHSSRLSINELLQFCSDPALLIASSPALVDQRT	388
		BWCoV	YIDGVPISTTGYHSELGVLNKNDSMSFSKMSIGELMRFAADPSLLVSASDAFVDLRT	387
	β	IBV	FVDGVVFFIATCGYHSELGIVMNDVNTMSFSKMGSLMQFVGDPAALLVGTSTNKLVDLRT	402
HCoV_OC43		FVDGVVFFVVSIGYHYKELGIVMNDVDTHRYRSLKDLLLYAADPALHVASASALYDLRT	389	
HCoV_HKU1		FVDGVVFFVVSIGYHYKELGVVNMNDVDTHRYRSLKDLLLYAADPAMHVASASALLDLRT	389	
MHV		FVDGVVFFVVSIGYHYKELGVVNMNDVDTHRYRSLKDLLLYAADPALHVASASALLDLRT	389	
MERS		FVDGVVFFVVSIGYHYKELGVVNMNDVSLHRHRLSLKELMMYAADPAMHIASSNAFLDLRT	394	
HKU4		FVDGVVFFIVSCGYHYKELGIVMNDVFNHRLALKELMMYAADPAMHIASSALWDLRT	395	
HKU5		FVDGVVFFIVSCGYHYKELGIVMNDVSLHRHRLSLKELMMYAADPAMHIASSALWDLRT	395	
HKU9		YVDGVVFFVSTGYHYRELGVVNMNDVVRQHAQRLSLRELLVYAADPAMHVAASNALSKRT	393	
SARS-CoV-2		FVDGVVFFVSTGYHFRRELGVVNMNDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRT	393	
SARS-CoV		FVDGVVFFVSTGYHFRRELGVVNMNDVNLHSSRLSFKELLVYAADPAMHAASGNLLLDKRT	393	
<p>*** .: : * * :***: * * .: : *: **: . . * **</p>				
δ	HKU19	PCQTIASIASGATKQSVKPGHFNAHFYEHASEGILSEDSGIDIRHFYMQDGEEAIAKDY	462	
	HKU11	PCQTLATMSSGITKQSVKPGHFNQHFYKHLSDILN-QLGIDLKHFFYMQDGEEAITDY	453	
	PDCV	PCQTLATMSSGIKQSVKPGHFNQHFYKHLSDNLNLL-QLGIDIRHFYMQDGEEAITDY	453	
	α	FCoV_65F	VCFSVAALSTGVTYQTVKPGHFNKDFYDFITERGFFEEGSELTLKHFFFAQGGEEAMTDF	450
		HCoV_229E	VCFSVAALSTGVTYQTVKPGHFNKDFYDFLRSQGFDEGSELTLKHFFFTQKGDAAIKDF	448
	α	HCoV_NL63	ICFSVAALSTGLTNQVVKPGHFNEEFYNFLRLRGFFDEGSELTLKHFFFAQNGDAAVKDF	448
		HKU8	VCFSVAALGTGLTKQTVKPGHFNKDFYDFLCAQGFDEGSELTLKHFFFAQKGDAAIRD	448
	γ	PEDV	VCFSVAALGTGMTNQTVKPGHFNKDFYDFLLEQGFDEGSELTLKHFFFAQKGDAAVKDF	448
		BWCoV	SCFSLSALSTGLTYQTVKPGHFNEDFYNFAEKKGFKEGSS IPLKHFFYIQDGNAAIADF	447
	β	IBV	SCFSVALASGITHQTVKPGHFNKDFYDFAEKAGMFKEGSS IPLKHFFYPQTGNAAINDY	462
HCoV_OC43		CCFSVAAITSGVKFQTVKPGHFNQDFYDFVLSKGLLKEGSSVDLKHFFFTQDGNAAITDY	449	
HCoV_HKU1		CCFSVAAITSGIKFQTVKPGHFNQDFYDFVKSGLFKEGSVTDLKHFFFTQDGNAAITDY	449	
MHV		CCFSVAAITSGVKFQTVKPGHFNQDFYDFVLSKGLLKEGSSVDLKHFFFTQDGNAAITDY	449	
MERS		ICFSVAALTTGLTFQTVRPGHFNKDFYDFVVSIGFFKEGSSVTLKHFFFAQDGNAAITDY	454	
HKU4		PCFSVAALTTGLTFQTVRPGHFNKDFYDFVVSIGFFKEGSSVTLKHFFFAQDGHAAITDY	455	
HKU5		PCFSVAALTTGLTFQTVRPGHFNKDFYDFVVSIGFFKEGSSVTLRHFFFAQDGHAAITDY	455	
HKU9		VCMSVAAMTTGVTQTVKPGQFNEEFYDFVFAIKCGFFKEGSTISFKHFFFAQDGNAAISDY	453	
SARS-CoV-2		TCFSVAALTNNVAFQTVKPGHFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDY	453	
SARS-CoV		TCFSVAALTNNVAFQTVKPGHFNKDFYDFAVSKGFFKEGSSVELKHFFFAQDGNAAISDY	453	
<p>* :.: : . . * * :***: * * . . .: : : :***: * * .***: *</p>				
δ	HKU19	SYRYRNTPTMVDIKQFLFVMEVADKYLSPYDGGCIPAEVTVVNNLDKSAGYPPNKLKGR	522	
	HKU11	SYRYRNTPTMVDIKMFLFVLEVADKYLQPYEGGCLNAQSVVNNLDKYAGYPPNKLKGR	513	
	PDCV	SYRYRNTPTMVDIKMFLFCLEVADKYLEPEYEGGCINAQSVVNNLDKSAGYPPNKLKGR	513	
	α	FCoV_65F	NYYRYNRVTVLDICQAQFVYKIVCKYFDCYDGGCINAREVVVNTNYDKSAGYPLNKFGR	510
		HCoV_229E	DYYRYNRPTILDIGQARVAYQVAARYFDCYEGGCITSREVVVNTLNKNSAGWPLNKFGR	508
	α	HCoV_NL63	DFYRYNKPTILDICQARVYTIKVSRYFDIYEGGCIAKACEVVVNTLNKNSAGWPLNKFGR	508
		HKU8	DFYRYNRPTVLDICQARVAYHVVKRYFDIYEGGCIAARDVVVNTLNKNSAGYPLNKFGR	508
	γ	PEDV	DYYRYNRPTVLDICQARVVYQIVQRYFDIYEGGCITAKEVVVNTLNKNSAGYPLNKFGR	508
		BWCoV	DYYRYNRPTMVDIQQLFCFEVTDKYFEYDGGCIPANQVVVNTLNKNSAGYPPNKFGR	507
	β	IBV	DYYRYNRPTMVDIQQLFCLEVTSKYFEYEGGCIPASQVVVNNLDKSAGYPPNKFGR	522
HCoV_OC43		NYYKYNLPTMVDIKQLLFVLEVYKYFEIYDGGCIPASQVI VNNYDKSAGYPPNKFGR	509	
HCoV_HKU1		NYYKYNLPTMVDIKQLLFVLEVYKYFEIYDGGCIPASQVI VNNYDKSAGYPPNKFGR	509	
MHV		NYYKYNLPTMVDIKQLLFVLEVYKYFEIYEGGCIPATQVI VNNYDKSAGYPPNKFGR	509	
MERS		NYYSYNLPTMCDIKQMLFCMEVVNKYFEIYDGGCLNASEVVVNNLDKSAGHPPNKFGR	514	
HKU4		SYAYNLPTMVDIKQMLFCMEVVNDKYFDIYDGGCLNASEVI VNNLDKSAGHPPNKFGR	515	
HKU5		SYAYNLPTMCDIKQMLFCMEVVNDKYFEIYDGGCLNASEVI VNNLDKSAGHPPNKFGR	515	
HKU9		DYYRYNLPTMCDIKQLLFSLEVVDKYFDYDGGCLQASQVVVANVDYDKSAGFPNKFGR	513	
SARS-CoV-2		DYYRYNLPTMCDIRQLLFVVEVVDKYFDYDGGCINANQVI VNNLDKSAGFPNKFGR	513	
SARS-CoV		DYYRYNLPTMCDIRQLLFVVEVVDKYFDYDGGCINANQVI VNNLDKSAGFPNKFGR	513	
<p>..* :* * : * * . . .: : .***: * * : * * * * * * * * *</p>				

SARS-CoV-2 E431R

Polymerase Motif G

δ	HKU19	NYVD-LTYAEQNAMFEYTKRNVLPVLTQMNLYAISAKDRARTVAGVSIISTMTNRQYHQ	581	Polymerase Motif F	
	HKU11	NYVD-MTYAEQNQLFEYTKRNVLPVLTQMNLYAISAKDRARTVAGVSIISTMTNRQYHQ	572		
α	PDCV	NYVD-MTHAEQNQLFEYTKRNVLPVLTQMNLYAISAKDRARTVAGVSIISTMTNRQYHQ	572		
	FCoV_65F	LYYETLSYEEQDAIFALTQRNVLPVLTQMNLYAISGKERARTVGGVSLSTMTTRQYHQ	570		
α	HCoV_229E	LYYESISYEEQDAIFLLTKRNVLPVLTQMNLYAISGKERARTVGGVSLSTMTTRQYHQ	568		
	HCoV_NL63	LYYESISYEEQDALFALTKRNVLPVLTQMNLYAISGKERARTVGGVSLSTMTTRQYHQ	568		
γ	HKU8	LYYESLSYEEQDALYALTKRNVLPVLTQMNLYAISGKERARTVGGVSLSTMTTRQYHQ	568		
	PEDV	LYYESLSYEEQDELAYTKRNVLPVLTQMNLYAISGKERARTVGGVSLSTMTTRQYHQ	568		
γ	BWCoV	LYYESLSYAEQDQLFELTKRNVLPVLTQMNLYAISAKSRARTVAGVSISSMTNRQYHQ	567		
	IBV	LYYE-MSLEEQDQLFESTKKNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	581		
β	HCoV_OC43	LYYEALSFEEQDEIYAYTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTGRMFHQ	569		
	HCoV_HKU1	LYYEALSFEEQNEIYAYTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTGRMFHQ	569		
β	MHV	LYYEALSFEEQDEIYAYTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTGRMFHQ	569		
	MERS	VYYESMSYQEDELFAVTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	574		
β	HKU4	VYYESMSYQEDELFAVTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	575		
	HKU5	VYYESLSYQEDELFAVTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	575		
β	HKU9	LYYESLSYADQDELFAVTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	573		
	SARS-CoV-2	LYYDSMSYEDQDALFAYTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	573		
β	SARS-CoV	LYYDSMSYEDQDALFAYTKRNVLPVLTQMNLYAISAKNRARTVAGVSISSMTNRQYHQ	573		
		: :: **: :: **:*::**:*::**. * *****.***: ::** * : **			
δ	HKU19	KLLKSISVARNQTIIVIGTTKFGGWDNMLRNLNMANINNPKLAGWDYPKCDRSPNLLRIT	641		Polymerase Motif A
	HKU11	KMLKSISLARNQTIIVIGTTKFGGWDNMLRRLMNGINNPILVGDYPKCDRSPNMLRIA	632		
α	PDCV	KMLKSISLARNQTIIVIGTTKFGGWDNMLRRLMNCINNPILVGDYPKCDRSPNMLRIA	632		
	FCoV_65F	KHLKLSIAATRNATVIVIGTTKFGGWDNMLKMLMRDNDGCLMGWDYPKCDRALPNMIRMA	630		
α	HCoV_229E	KCLKSIVATRNATVIVIGTTKFGGWDNMLKMLMADVDDPKLMGWDYPKCDRAMPNIRML	628		
	HCoV_NL63	KHLKLSIVNTRNATVIVIGTTKFGGWNMLRRTLIDGVENPMLMGWDYPKCDRALPNMIRMI	628		
γ	HKU8	KHLKLSIVNTRNATVIVIGTTKFGGWDNMLRNLMDGVNACLMDGWDYPKCDRALPNMIRMI	628		
	PEDV	KHLKLSIVNTRGASVIVIGTTKFGGWDNMLKMLIDGVENPCLMGWDYPKCDRALPNMIRMI	628		
γ	BWCoV	KCLKSIVNTRNATVIVIGTTKFGGWDNMLRNLMDGVNACLMDGWDYPKCDRALPNMIRMI	627		
	IBV	KILKSIVNTRNAPVIVIGTTKFGGWDNMLRNLIQGVEDPILMGWDYPKCDRAMPNLLRIA	641		
β	HCoV_OC43	KCLKSIAATRGPVIVIGTTKFGGWDMLRRLIKDVDNPLVLMGWDYPKCDRAMPNLLRIV	629		
	HCoV_HKU1	KCLKSIAATRGPVIVIGTTKFGGWDMLRRLIKDVDNPLVLMGWDYPKCDRAMPNILLRIV	629		
β	MHV	KCLKSIAATRGPVIVIGTTKFGGWDMLRRLIKDVDSPVLMGWDYPKCDRAMPNILLRIV	629		
	MERS	KMLKSMAATRGCATVIGTTKFGGWDVFLKTLTKYKVDNPHLMGWDYPKCDRAMPNMCRI	634		
β	HKU4	KMLKSMAATRGCATVIGTTKFGGWDVFLKTLTKYKVDNPHLMGWDYPKCDRAMPNMCRI	635		
	HKU5	KMLKSMAATRGCATVIGTTKFGGWDVFLKTLTKYKVDNPHLMGWDYPKCDRAMPNMCRI	635		
β	HKU9	KMLKSIAAARGASVIVIGTTKFGGWDNMLRNLRTLCGVDNPHLMGWDYPKCDRAMPNLLRIF	633		
	SARS-CoV-2	KLLKSIAATRGCATVIVIGTSKFGGWHNMLKTVSDVENPHLMGWDYPKCDRAMPNMLRIM	633		
β	SARS-CoV	KLLKSIAATRGCATVIVIGTSKFGGWHNMLKTVSDVETPHLMGWDYPKCDRAMPNMLRIM	633		
		* ***: :*. *****. **: : .: : * *****:*. : *			
δ	HKU19	SSLLLARKH-ACCTHSQRFYRLANCAQVLSEIVVSGNVMYVKPGGTSSSGDATTAYANSV	700	Polymerase Motif B	
	HKU11	ASCLLARKH-TCCNQSQRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	691		
α	PDCV	ASCLLARKH-TCCNQSQRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	691		
	FCoV_65F	SAMVLGSKHIGCCTHSRDRFYRLSNELAQVLTEVVVHCTGGFYIKPGGTTSGDGTAYANSA	690		
α	HCoV_229E	SAMVLGSKHVTCTASDKFYRLSNELAQVLTEVVVHCTGGFYIKPGGTTSGDATTAYANSV	688		
	HCoV_NL63	SAMVLGSKHVNCCTATDRFYRLGNELAQVLTEVVVHCTGGFYIKPGGTTSGDATTAYANSI	688		
γ	HKU8	SAMVLGSKHVNCCTNSDRFYRLCNELAQVLTEVVVHCTGGFYIKPGGTTSGDATTAYANSV	688		
	PEDV	SAMVLGSKHTTCCSSTRDRFYRLCNELAQVLTEVVVHCTGGFYIKPGGTTSGDATTAYANSV	688		
γ	BWCoV	ASLILARRHKCCDWNERIYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSA	687		
	IBV	ASLVLARKHTNCTWSEVRYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	701		
β	HCoV_OC43	SSLVLARKHETCCSQRDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAFANSV	689		
	HCoV_HKU1	SSLVLARKHETCCSQRDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAFANSV	689		
β	MHV	SSLVLARKHDSCHTDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAFANSV	689		
	MERS	ASLILARKHGTCTTDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	694		
β	HKU4	ASLILARKHSTCTNSDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	695		
	HKU5	ASLILARKHSTCTNSDRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	695		
β	HKU9	ASLILARKHSTCCNASERFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDSTAYANSV	693		
	SARS-CoV-2	ASLVLARKHTTCCSLSHRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	693		
β	SARS-CoV	ASLVLARKHNTCCNLSHRFYRLANCAQVLSEIVVSGNLYVVKPGGTSSSGDATTAYANSV	693		
		:: :*. :* ** .: :** ** .***:* . . . * .*****:***:***:***			

δ	HKU19	FNILQVVSANIRFMSTSAATHHDVDMHLHRQIYDDIYRGSNSDSVAIQSFYEHQKYF	760	
	HKU11	FNILQVVSANVATFLSTSTSSHSNREIADLHRNLYEDIYRGSNNTTIIIDQFYQHLQKYF	751	
	PDCV	FNILQVVSANVATFLSTSTTTHLNKDIADLHRSLYEDIYRGSNDITVINRFYQHLQSYF	751	
	α	FCoV_65F	FNIFQAVSANVNKLLGVDSTNCNNVTVKSIRKRIYDNCYRSSSVDDDFVVEYFSLRKHF	750
		HCoV_229E	FNIFQAVSSNINCVLSVSSNCCNNFNVKKLQRLYDNCYRNSNVDSEFVDDFYGYLQKHF	748
	α	HCoV_NL63	FNIFQAVSSNINRLLSVPDSCNNVNRDLQRRLYDNCYRLTSVEESFIDDYGYLQKHF	748
		HKU8	FNIFQAVSANINRILGINSNTCNNLAVKSLQRMLYDNCYRSSAVDPGFVDTFYGYLRKHF	748
	γ	PEDV	FNIFQAVSANVNKLLSVDNVCNLEVKLQRLYECCYRSTTVDDQFVVEYGYLRKHF	748
		BWCoV	FNLFQATAANVAQLLATPSTRIFYAEVRLQHELYTQVYRRDKPDMDFVYTFYAYLNKHF	747
	β	IBV	FNIIQATSANVARLLSVITRDIYVDDIKSLQYELYQVYRRVNFDPAFVKEFYSYLCKNF	761
HCoV_OC43		FNICQAVSANVCALMSCNGNKIEDLSIRALQKRLYSHVYRSDKVDSTFVTEYEFNLKHF	749	

Polymerase
Motif B
Motif C

δ	HKU19	GLMILSDDGVACIDQEAAKQGMVADLDDFRDVLFYQNNVYMSDSKCIWETDMSKGPHEFC	820	
	HKU11	GLMILSDDGVACIDTEAAASGVVSNLDGFRDILFYQNNVYMSDSKCIWETDMDTVGPHFC	811	
	PDCV	GLMILSDDGVACIDSAVAKAGAVADLDGFRDILFYQNNVYMSDSKCIWETDMDNVGPHFC	811	
	α	FCoV_65F	SMMILSDDGVVVCYNKDYADLGYVADISAFKATLYYQNNVFMSTAKCWVEPDLNVGPHFC	810
		HCoV_229E	SMMILSDDGVVVCYNKTYAELGYIADISAFKATLYYQNGVFMSTAKCWTEEDLSIGPHFC	808
	α	HCoV_NL63	SMMILSDDGVVVCYNKDYAELGYIADISAFKATLYYQNNVFMSTAKCWVEEDLTGKPHFC	808
		HKU8	SMMILSDDGVVVCYNKEYASLGYVADINAFKATLYYQNNVFMSTAKCWVEEDLTGKPHFC	808
	γ	PEDV	SMMILSDDGVVVCYNNDYASLGYVADLNAFKAVLYYQNNVFMSTAKCWVEEDLTGKPHFC	808
		BWCoV	SLMILSDDGVVVCYNKSYAEGMVASIASFREVLFYQNNVFMSTAKCWTEEDVKGPHFC	807
	β	IBV	SLMILSDDGVVVCYNNTLAKQGLVADISGFREVLYYQNNVFMSTAKCWVEEDLTGKPHFC	821
HCoV_OC43		SMMILSDDGVVVCYNSDYASKGYIANISAFQVLYYQNNVFMSESKCWVEHDINNGPHFC	809	

Polymerase
Motif C
Motif D
Motif E

SARS-CoV-2
D760A, A797R,
E802R, H810R

δ	HKU19	SQHTVLAHYDGEPCYYPYDPVSRILGACIFVNETEKTPVQNLERYISLAIDAYPLTKVD	880	
	HKU11	SQHTVLAHEHGKPYLPPYDPVSRILGACIFVDDVKNADPQNLERYISLAIDAYPLTKVD	871	
	PDCV	SQHTVLAHEHDGKPYLPPYDPVSRILGACIFVDDVKNADPQNLERYISLAIDAYPLTKVD	871	
	α	FCoV_65F	SQHTLQIVGADGDYLYLPPYDPSTRILSAGVFVDDIVKTDNVIMLERYVSLAIDAYPLTKHP	870
		HCoV_229E	SQHTMQIVDENGKYYLPPYDPSTRISAGVFVDDVTKTDAVILLERYVSLAIDAYPLSKHP	868
	α	HCoV_NL63	SQHTMQIVDKDGTYYLPPYDPSTRILSAGVFVDDVVKTDVAVLLERYVSLAIDAYPLSKHP	868
		HKU8	SQHTMQIVDGDGTYYLPPYDPSTRILSAGVFVDDVVKTDVAVLLERYVSLAIDAYPLSKHP	868
	γ	PEDV	SQHTMQIVDKDGTYYLPPYDPSTRILSAGVFVDDVVKTDVAVLLERYVSLAIDAYPLSKHE	868
		BWCoV	SQHSMLVEIDGEMRYLPPYDPSTRILGACVFVDDVKEPVMERYVALAIDAYPLIYHE	867
	β	IBV	SQHTMLVEVDGEPYRYPYDPSTRILCACVFVDDLDKTESVAVMERYIALAIDAYPLVHHE	881
HCoV_OC43		SQHTMLVKMDGDDVYLYLPPYDPSTRILGACVFVDDLLKTDVLLIERFVSLAIDAYPLVYHE	869	

Polymerase
Motif E

SARS-CoV-2 H816R

δ	HKU19	-NKKGKVFYVLLDYIRKLANELQEGIMDAFQSTSDTSTYINNFTENFYSDMYAKAPVLQ	938
	HKU11	-PIKGVFYVLLLDYIRILAQELQDGIIDTFQSMTDMSYVNNFVQEAFFYAQMIEQSPVTLQ	929
	PDCV	-PIKGVFYVLLLDYIRVLAQELQDGIIDAFQSLTDMSYVNNFMNEAFYAQMIEQSPVTLQ	929
α	FCoV_65F	KPAYQKVFYALLDQVVKHLQKTLNAGILDSFSVTMLLEDGQDKFWSEEFYASLYEKSTVLQ	929
	HCoV_229E	KPEYRKVFYALLDQVVKHLNKTLLNEGVLSEFSVTLLDEQESKFWDESFYASMYEKSTVLQ	927
	HCoV_NL63	NSEYRKVFYVLLDQVVKHLNKNLNEGVLSEFSVTLLDNQEDKFWCEDFYASMYENSTVLQ	927
	HKU8	NPEYRKVFYVLLDQVVKHLNNTLNQGVLEFSVTLLDASSKFWDESFYANLYEKSAVLQ	927
	PEDV	NPEYKVFYVLLDQVVKHLYKTLNAGVLSEFSVTLLDSTAKFWDESFYANMYEKSAVLQ	927
γ	BWCoV	NEEYKGVFYVLLSYIQTLYQRLSNDMLMDYSFVMDIDKGSKFWEQEFYENMYRAPVTLQ	926
	IBV	NEEYKVFVLLSYIRKLYQELSQNMLMDYSFVMDIDKGSKFWEQEFYENMYRAPVTLQ	940
β	HCoV_OC43	NEEYQVFRVYLAYIKKLYNDLGNQILDSYSVILSTCDGQKFTDESFYKNMYLRSVAVMQ	928
	HCoV_HKU1	NEEYQVFRVYLEYIKKLYNDLGTQILDSYSVILSTCDGLKFTDESFYKNMYLRSVAVMQ	928
	MHV	NPEYQVFRVYLEYIKKLYNDLGNQILDSYSVILSTCDGQKFTDETFYKNMYLRSVAVLQ	928
	MERS	DIYQNVFVYLYQYIEKLYKDLTGHLMDYSVMLCGDNSAKFWEEAFYRDLYSSPTTLQ	933
	HKU4	DTEYQNVFVYLYQYIEKLYKDLTGHLMDYSVMLCGDSDAKFWEEGFYRDLYSSPTTLQ	934
	HKU5	DPEYQNVFVYLYQYIEKLYKDLTGHLMDYSVMLCGDNSAKFWEEGFYRDLYTAPTTLQ	934
	HKU9	DPEYQNVFVYLYQYIEKLYKDLTGHLMDYSVMLASDNASKYWEVDFYENMYMESATLQ	932
	SARS-CoV-2	NQYADVFHLYLYQYIRKLHDELGTGHLMDYSVMLTNDNTSRWEPFYEAMYPHTVTLQ	932
	SARS-CoV	NQYADVFHLYLYQYIRKLHDELGTGHLMDYSVMLTNDNTSRWEPFYEAMYPHTVTLQ	932

.** * :. * . * :: :. :. :. ** :* :*

Figure S8, sequence alignment of coronavirus nsp12s: Global alignment of nsp12 from *AlphaCoV* (FCoV-65F, HCoV-229E, HCoV-NL63, HKU8, PEDV), *BetaCoV* (HCoV-OC43, HKU1, MHV, MERS, HKU4, HKU5, HKU9, SARS-CoV-2, and SARS-CoV), *GammaCoV* (BWCoV, IBV), and *DeltaCoV* (HKU19, HKU11, PDCV) genera. Global alignment was done using Clustal Omega. SARS-CoV-2 residues mutated during this study are highlighted with a red box and listed to the right of the sequences. Residues marked with “*” are conserved, “:” are very similar, and “.” are moderately similar residues.

**Chapter 5: Structural insight into araCTP inhibition of the SARS-CoV-2
polymerase complex**

Authors: Thomas K. Anderson¹, Jamie J. Arnold², Craig E. Cameron², Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

²Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA.

Author contributions:

T.K.A. generated all data and figures presented in this chapter.

T.K.A., R.N.K., J.J.A., and C.E.C equally contributed to the development of this project.

T.K.A. wrote this chapter with input and editing by R.N.K.

Abstract:

Coronaviruses are a subfamily of viruses that cause a wide range of disease severities in both humans and animals. Since 2002, three coronaviruses have crossed over from animal reservoirs into humans causing epidemics or pandemics. Currently there are only three drugs approved for the treatment of coronavirus infection, with the usefulness of two under debate. As coronaviruses continue to pose a threat to global health, we will require additional antivirals to treat coronavirus induced diseases and prevent future pandemics. Arabinose nucleosides are a group of nucleoside analogues that have been used for the treatment of herpes virus infections and various types of leukemia. Here, we show that the arabinose nucleotide araCTP is a viable substrate for the SARS-CoV-2 RNA-dependent RNA polymerase, and that its incorporation terminates RNA elongation. Using single-particle cryoEM, we show that araCTP has multiple possible mechanisms for elongation termination. Therefore, arabinose nucleotides are attractive options for the development of broadly acting coronavirus antivirals to treat current coronavirus diseases and help prepare us for the next coronavirus spillover.

Introduction:

Since the turn of the century, three coronaviruses (CoVs) have crossed over from animal reservoirs into humans causing epidemics or pandemics. SARS-CoV emerged in 2002 and caused over 8,000 cases resulting in >800 deaths (1,2). In 2012, MERS-CoV emerged and continues to persist through sporadic outbreaks with reported mortality rates as high as 35% (3-5). The emergence of SARS-CoV-2 led to the ongoing COVID-19 pandemic which has caused over 7 million deaths globally since 2019 (6,7). Although the impact of COVID-19 has been extreme, its emergence also drove the rapid development of several vaccines. While the vaccines have proven effective, their long-term efficacy is a concern as SARS-CoV-2 continues to evolve into new variants (8). The continuing trend of CoV crossovers and new SARS-CoV-2 variants highlight the ever-present CoV threat to human health. Currently there are only three approved antiviral treatments for CoV induced diseases in humans: Remdesivir, Molnupiravir, and Paxlovid (9). To limit the pandemic potential of future CoV outbreaks, we need additional broadly acting and highly effective CoV antivirals.

CoVs are enveloped RNA viruses with large 30 kb (+) single-stranded RNA (ssRNA) genomes with a 5' cap and 3' poly-adenosine tail (10,11). The 5' two-thirds of the CoV genome has two large open reading frames (ORF1a and ORF1b) that encode for two polyproteins (12). During translation, a -1 ribosomal frameshift at the end of ORF1a allows translation of ORF1b to produce a second, larger polyprotein (13,14). The two polyproteins are cleaved by viral proteases non-structural protein 3 (nsp3) and nsp5 into 11 or 15 nsp subunits (10). Three nsps localize to the host endoplasmic reticulum (ER) where they remodel the ER membrane into a network of double-membrane vesicles

(DMVs) (15-17). Within DMVs the majority of nsps assemble into the viral replication transcription complex (RTC) and synthesize viral RNA products (16,18). The central component of the RTC is the viral RNA-dependent RNA polymerase (RdRP) nsp12, which with protein replication cofactors nsp7 and nsp8, forms the minimum complex needed to synthesize RNA in a processive manner; we refer to this complex as the core-RTC (19,20). The nsp12 polymerase has a secondary active site termed the nidovirus RdRP-associated nucleotidyltransferase (NiRAN) that has been implicated in RNA capping (21-23). At 30 kb, CoVs have some of the largest known RNA virus genomes. To maintain their genomic integrity CoVs encode their own proofreading enzyme in nsp14, a 3'-5' exoribonuclease (24-27). The first structure of a CoV core-RTC revealed that each nsp12 binds one nsp8, and one nsp7+nsp8 heterodimer (21). A structure of the SARS-CoV-2 core-RTC bound to a dsRNA substrate revealed that each of the two nsp8s have long helical extensions that bind dsRNA leaving the active site (28). This interaction was predicted to promote the processive replication of the core-RTC, but more recent work on *Beta-* and *Alphacoronavirus* RTCs indicates that only one of the N-terminal helices of the nsp8s is needed for RNA binding and processive polymerase activity (19,29).

Prior work on (+) ssRNA virus RdRPs, in particular those of picornaviruses and hepatitis C virus, have shown that they share a conserved structure composed of seven motifs, A-G, that form a shape that has been described as a cupped right hand with a fingers, palm, and thumb domain with the RdRP active site in the palm (30-34). The “fingertips” of the RdRP’s fingers domain reach over the active site to contact the thumb domain and enclose the active site (31). This interaction creates two channels within the

polymerase: one for bound RNA and the other for NTP entry (31). Structures of RdRPs bound to RNA revealed that prior to NTP binding, the +1 templating RNA is base stacked with the template base -1 upstream of it in the RdRP active site (35). This differs from DNA polymerases that have separate NTP binding sites (36). Upon NTP binding, the RdRP active site closes, most notably by a shift in motif A (33,35). Active site closure positions active site residues and two Mg^{2+} ions to coordinate catalysis of the primer RNA 3'OH's nucleophilic attack of the incoming NTP's α -phosphate. After nucleophilic attack, motif A shifts away re-opening the active site and allowing release of the pyrophosphate product (33). RNA then moves through the polymerase active site by a Brownian ratchet mechanism, poising downstream template RNA for NTP binding and further elongation (33). More recent work on the SARS-CoV-2 core-RTC has shown that nsp12 uses homologous mechanisms for NTP binding and active site closure (37).

Nucleoside analogues (NA) are a common type of antiviral that mimic NTPs and dNTPs to be incorporated by viral polymerases (38,39). After incorporation, NAs block virus replication by terminating RNA elongation or inducing hyper-mutagenesis. Ribavirin, a guanosine analogue, is a broadly acting and well characterized NA that inhibits viruses such as hepatitis C virus and Lassa fever virus replication by causing hyper-mutagenesis or inhibiting the host inosine-5'-monophosphate dehydrogenase (IMPDH) which limits GTP availability (40-43). Sofosbuvir is a highly effective hepatitis C virus NA that terminates RNA synthesis after its incorporation (44). CoVs' ability to proofread during replication makes commonly used NAs such as ribavirin and 5-fluorouracil ineffective in blocking replication (45-47). In tissue culture models it has been shown that knocking out CoVs' proofreading ability renders them susceptible to NAs such as 5-fluorouracil (47).

After the emergence of SARS-CoV-2, two NAs were approved for emergency use treatment of COVID-19. The NA Remdesivir, an adenosine mimetic with a 1' cyano substitution, was able to block SARS-CoV-2 replication *in vitro* and *in vivo* (48-50). Remdesivir showed promise in treating COVID-19 in clinical trials but was later shown to be ineffective at reducing symptoms and hospitalizations and the World Health Organization has recommended against its use for treating human CoV infections (51-53). Molnupiravir, a hypermutator cytosine mimetic, was shown to be a potent CoV antiviral in tissue culture and animal models (54,55). Clinical trials showed Molnupiravir was effective at treating COVID-19 in patients and in 2022 it received emergency use authorization for treatment (56). Although promising, the efficacy of Molnupiravir is debated due to its inability to reduce COVID-19 hospitalizations or death among high-risk adults (57). Structural studies of the SARS-CoV-2 core-RdRp interacting with nucleoside analogues have provided important information into their mechanisms of action. Structures of Remdesivir in pre-incorporated and incorporated states revealed that its 1' cyano substitution sterically clashes with nsp12 residue Ser861 in the RdRp active site (37,58). This clash stalls the translocation of Remdesivir-incorporated nascent RNA from the +3 to +4 position. Structures of Molnupiravir in nsp12's active site base paired with either adenosine or guanosine have demonstrated its ability to induce hyper-mutagenesis by being a viable partner for base-pairing with multiple nucleotides in the RdRp active site (59).

Arabinose nucleotides (araNTPs) were some of the first NAs to be identified as having medicinal uses (60). AraNTPs are nucleotides that have flipped chirality at the ribose 2'OH (**Figure 1A**) which alters the ribose's sugar pucker from C3' endo to C2' endo

(61). In 1976 araATP was approved to treat herpes simplex virus and varicella zoster virus infections but was eventually replaced by the more effective and less toxic acyclovir (62,63). AraCTP and 2-fluoro-araATP are two araNTPs that are still commonly used to treat different types of leukemia (62,64-68). To our knowledge arabinose nucleotides and their derivatives have not been tested as CoV antivirals and could be used to develop new CoV treatments.

In this chapter, I present a structure of an elongating SARS-CoV-2 core-RTC that has been stalled by araCTP. I have identified that araCTP has several possible mechanisms by which it may inhibit coronavirus RNA synthesis. This work demonstrates that arabinose nucleotides are substrates for coronavirus replication complexes and could be used as a scaffold to develop more effective and broadly acting antivirals.

Methods:

Recombinant Protein Expression:

Recombinant Protein Expression

SARS-CoV-2 nsp7 and nsp8 replication factors were expressed in Rosetta 2pLysS *E. coli* cells (Novagen). Cultures were grown at 37°C and at an OD₆₀₀ of 0.6-0.8 were induced with isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 500 μM. After shaking for 16 hours at 16°C, cells were harvested by centrifugation and resuspended in Ni-wash buffer (10 mM Tris-Cl pH 8.0, 300 mM NaCl, 30 mM imidazole, and 2 mM dithiothreitol (DTT)). Cells were lysed in a microfluidizer (Microfluidics) and lysates cleared via centrifugation, supernatants were further cleared by 0.45 μm filtration. Proteins were then purified using Ni-NTA agarose beads (Qiagen), eluting with the same buffer with 300 mM imidazole. Eluted protein was digested with TEV protease (1% w/w)

overnight at 4°C while dialyzing (10 mM Tris-Cl pH 8.0, 300 mM NaCl, and 2 mM DTT). Undigested protein was removed by flowing dialyzed protein back over Ni-NTA agarose beads. Digested protein was purified using a Superdex 200 Increase 10/300 GL column (Cytiva) in 25 mM Tris-Cl pH 8.0, 300 mM sodium NaCl, and 2 mM DTT. Fractions containing the protein of interest were concentrated using ultrafiltration. Concentrated protein was aliquoted, flash frozen in liquid nitrogen, and stored at -80°C.

SARS-CoV-2 nsp12 was cloned into a pFastBac plasmid and DH10Bac *E. coli* (Life Technologies) were used to produce a recombinant bacmid. Bacmids were transfected into Sf9 cells (Expression Systems) with Cellfectin II (Life Technologies) to produce recombinant baculoviruses. Sf9 cells were used to amplify the initial baculovirus stock twice. Amplified baculoviruses were used to infect Sf21 cells (Expression Systems) for protein expression. After two days of incubation and gentle shaking at 27°C, cells were collected via centrifugation and pellets resuspended in wash buffer (25 mM HEPES pH 7.4, 300 mM NaCl, 1 mM MgCl₂, and 5 mM DTT) and 143 µL of BioLock (IBA) per 1 L of culture. Cells were lysed using a microfluidizer (Microfluidics) and lysates cleared via centrifugation and supernatants were further cleared by 0.45 µm filtration. Protein was purified using Streptactin superflow agarose (IBA) and eluted with wash buffer that contained 2.5 mM desthiobiotin. Protein was further purified via size exclusion chromatography with a Superdex 200 Increase 10/300 GL column (Cytiva) in 25 mM HEPES pH 7.4, 300 mM NaCl, 100 µM MgCl₂, and 2 mM tris(2-carboxyethyl)phosphine (TCEP). Nsp12 containing fractions were pooled and concentrated using ultrafiltration. The protein was then aliquoted, flash frozen in liquid nitrogen and stored at -80°C.

RNA Substrate Design and Annealing:

RNA oligos were designed and purchased from Integrated DNA Technologies (IDT). The primer RNA had a 5' fluorescein (6-FAM) tag to monitor its extension. To anneal the RNA, oligoribonucleotides were mixed (with template in 20% excess to primer) in RNA annealing buffer (2.5 mM HEPES, pH 7.4, 2.5 mM KCl and 0.5 mM MgCl₂) and heated at 95°C for 5 minutes before slowly cooling back down to 25°C for 1.25 hours.

Primer RNA sequence:

5' - CAUUCUCCUAAGAAGCUAUUAAAUCACA - 3'

Template RNA sequence:

5' - AAAAAGGGUUGUGAUUUUAAUAGCUUCUUAGGAGAAUG - 3'

In vitro Primer Extension:

Duplex RNA was the limiting reagent for core-RTC formation at 0.25 μM with nsp12 at 0.5 μM and, nsp7 and nsp8 each at 1.5 μM. Reaction buffer conditions were 10 mM Tris-Cl pH 8.0, 10 mM NaCl, 2 mM MgCl₂, and 1 mM DTT. Proteins were mixed and incubated at 25°C for 15 minutes prior to the addition of RNA. After the RNA was added, reactions were incubated at 25°C for 15 minutes. Reactions were initiated with the addition of NTPs and/or araCTP and incubated at 25°C for 1 minute before the reaction was halted by addition of loading buffer (95% (w/v) formamide, 2 mM ethylenediaminetetraacetic acid (EDTA), and 0.75 mM bromophenol blue). Samples were heated at 95°C for 15 minutes before being analyzed by denaturing urea-PAGE (8 M urea, 15% polyacrylamide) run in 1X TBE (89 mM Tris-Cl, pH 8.3, 89 mM boric acid, 2 mM EDTA). Gels were imaged on a Typhoon FLA 9200 (GE Healthcare) using excitation and emission values of 470 and 530 nm, respectively.

CryoEM Data Collection, Processing, and Model Building:

For structure determination, the core-RTC was assembled at a total protein concentration of 2 mg/mL with a nsp7 : nsp8 : nsp12 : RNA ratio of 2 : 3 : 1 : 1.2 in cryoEM sample buffer (25 mM HEPES, pH 7.5, 100 mM NaCl, 2 mM MgCl₂, 2 mM DTT). Proteins were combined in sample buffer and incubated at 25°C for 15 minutes. Then RNA was added, and the sample incubated at 25°C for 15 minutes before being concentrated to ~4 mg/mL total protein concentration using 100 kDa MW cutoff ultrafiltration. Once concentrated, ATP and araCTP were added to the sample at a final concentration of 800 μM each and extension reactions were incubated at 25°C for 2 minutes before being put on ice until vitrification.

Samples were prepared for cryoEM on UltraAuFoil R1.2/1.3 300 mesh grids (Quantifoil) using a Vitrobot Mark IV (ThermoFisher Scientific). Grids were freshly glow discharged using a GloQube Plus (Quorum) for 20 seconds with a current of 20 mA using an air atmosphere to create a negative surface charge. Immediately before blotting, 0.5 μL of 42 mM 3-([3-Cholamidopropyl] dimethylammonio)-2-hydroxy-1-propanesulfonate (CHAPSO) was added to 3 μL of sample (final [CHAPSO] = 6 mM). Sample was then spotted onto grids and excess sample removed via double-sided blotting before plunge freezing in liquid ethane. The Vitrobot chamber was kept at 100% humidity and 4°C during the process.

Grids were screened and data collected using EPU on a Talos Arctica 200 kV transmission electron microscope (ThermoFisher Scientific). Movies were collected on a Gatan K3 direct electron detector in CDS mode with a GIF quantum energy filter slit width

of 20 eV. Data was collected at a magnification of 79,000x, a pixel size of 1.064 Å, and a defocus range of -0.5 to -2.0 µm with a step size of 0.5 µm.

Data was processed using cryoSPARC v4.4.0 (69). Following patch motion correction, and patch CTF estimation, particles were picked using blob picker and extracted with a box size of 256 pixels. Low quality picks were removed by one round of 2D classification. Remaining particles were then used for ab-initio reconstruction followed by heterogenous reconstruction and non-uniform refinement. Particles were then sorted with another round of heterogenous refinement and a final non-uniform refinement.

To build the SARS-CoV-2 araCTP model, we docked a SARS-CoV-2 model (PDB ID: 7UOE) into our electron density using ChimeraX and mutated the RNA sequence to match our RNA duplex in Coot (37,70,71). We validated the sequence of our RNA (and araCTP incorporation) by checking the order of purines and pyrimidines in the RdRP active site based on the size of electron density for the nitrogenous bases. The CTP ligands in this model were deleted and the protein + RNA model was initially refined and validated using ISOLDE, COOT, and PHENIX real space refinement (71-73). After this, araCMP (3 letter code CAR) and araCTP (3 letter code HF4) ligand models were placed into our map using COOT (71). Final refinements of our complete model were done using PHENIX real space refinement with the necessary ligand restraint files (73).

Results:

Assembly of an active SARS-CoV-2 Polymerase Complex:

To study the impact of araCTP on the SARS-CoV-2 core-RTC we recombinantly expressed and purified the viral replication cofactors nsp7 and nsp8 using a bacterial expression system and the viral RdRP nsp12 using a baculovirus-insect cell system

(**Figure 1B**). We then assembled the core-RTC and tested its RdRP activity on a short RNA duplex using an *in vitro* primer extension assay and saw polymerase activity in the presence of all three nsps (**Figure 1C and D**).

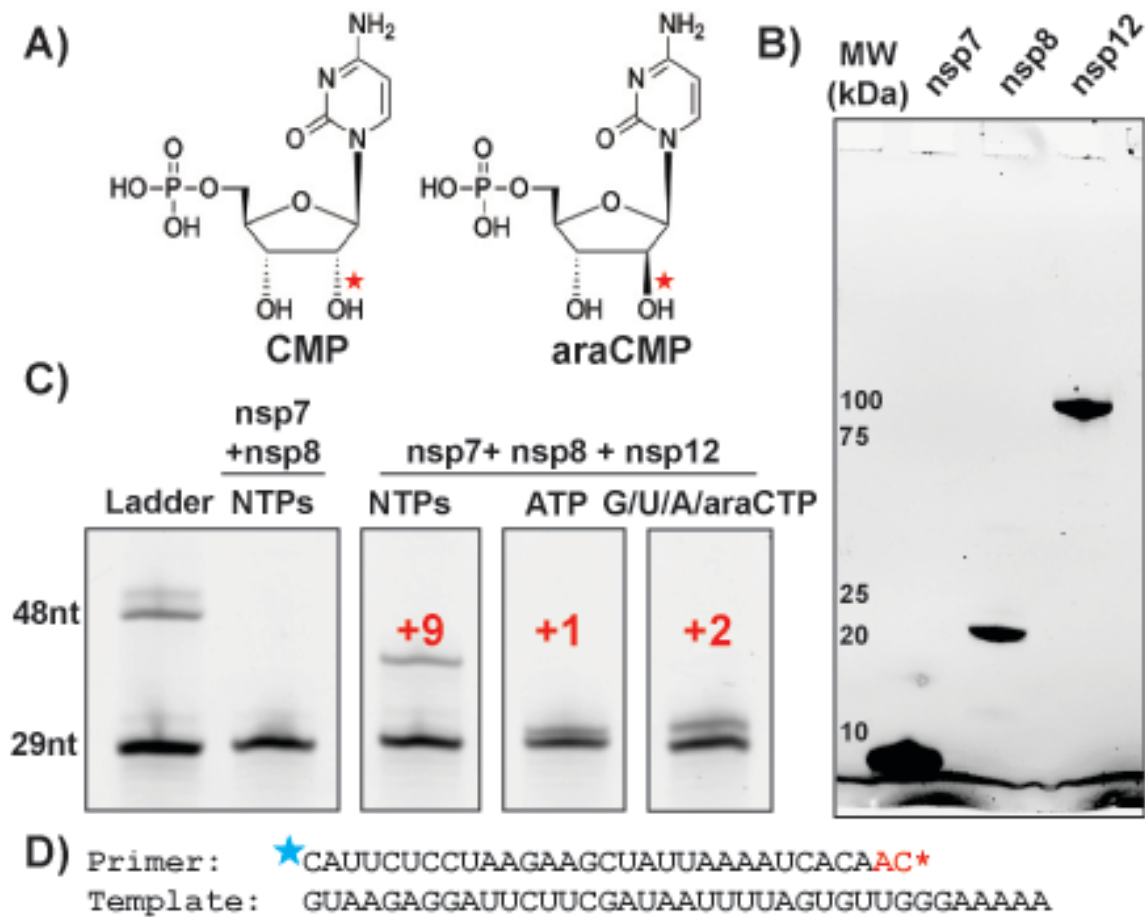


Figure 1, AraCTP inhibits the SARS-CoV-2 core-RTC: **A)** comparison of CMP and araCMP with the flipped chirality 2'OH marked by a red star. **B)** SDS-PAGE analysis of recombinant SARS-CoV-2 nsp7 (9,300 Da), nsp8 (21,900 Da), and nsp12(110,100), numbers indicate MW (kDa) positions. **C and D)** *In vitro* primer extension of a short RNA primer-template pair (**D**) in the presence of various NTPs and/or araCTP. The first two nucleotides to be incorporated during extension (ATP and CTP/araCTP (*)) are colored red (**D**).

AraCTP stalls SARS-CoV-2 polymerase elongation

Using our *in vitro* RNA primer extension assay we sought to determine if araCTP could be used as a substrate by the SARS-CoV-2 core-RTC, and if its incorporation would inhibit elongation. Extension in the presence of G/U/ATP + araCTP induced premature elongation termination after just two nucleotide incorporations (**Figure 1C**). Based on the template sequence, the two-nucleotides would be AMP-araCMP with no further araCMPs incorporated (**Figure 1D**). These results demonstrate that araCTP is a substrate for the SARS-CoV-2 core-RTC and that it terminates RNA extension.

Structure of an araCTP stalled SARS-CoV-2 polymerase complex

To solve the structure of a stalled SARS-CoV-2 core-RTC, we assembled the RTC and performed RNA extension using the same RNA duplex as above in the presence of ATP and araCTP. Following elongation, we vitrified our stalled complexes and solved the structure using single-particle cryoEM (**Figure 2, Table S1, Figure S1 and S2**). Our final cryoEM reconstruction has an atomic resolution of 3.3 Å and contains density for the majority of nsp12, nsp7, two protomers of nsp8, and two-turns of dsRNA exiting the polymerase active site (**Figure 2B**). The difference in size of purine and pyrimidine bases allowed us to confidently assign the sequence of our RNA primer and templates. Sequence assignment confirmed that an AMP and araCMP were sequentially incorporated at the 3' terminus (**Figure 2C**). In addition to the araCMP there was an unincorporated araCTP base-paired with the +1 template GMP in the RdRP active site (**Figure 2C**). The presence of an unincorporated araCTP in nsp12's active site emphasizes araCTP's ability to potently inhibit RNA elongation. The araCTP is coordinating a Mg²⁺ ion as is expected with common two-metal mechanisms of catalysis

but we lacked strong density for a second Mg^{2+} (**Figure 2C**) (74). The Mg^{2+} present in our structure is referred to as metal B in two-metal mechanisms while the Mg^{2+} absent is metal A. The NiRAN domain of nsp12 was well resolved with additional density for a nucleotide-triphosphate with a coordinated Mg^{2+} . Based off the size of electron density and nucleotides provided we built this density to be araCTP, although it should be noted that the density for the gamma-phosphate is weak (**Figure 2D**).

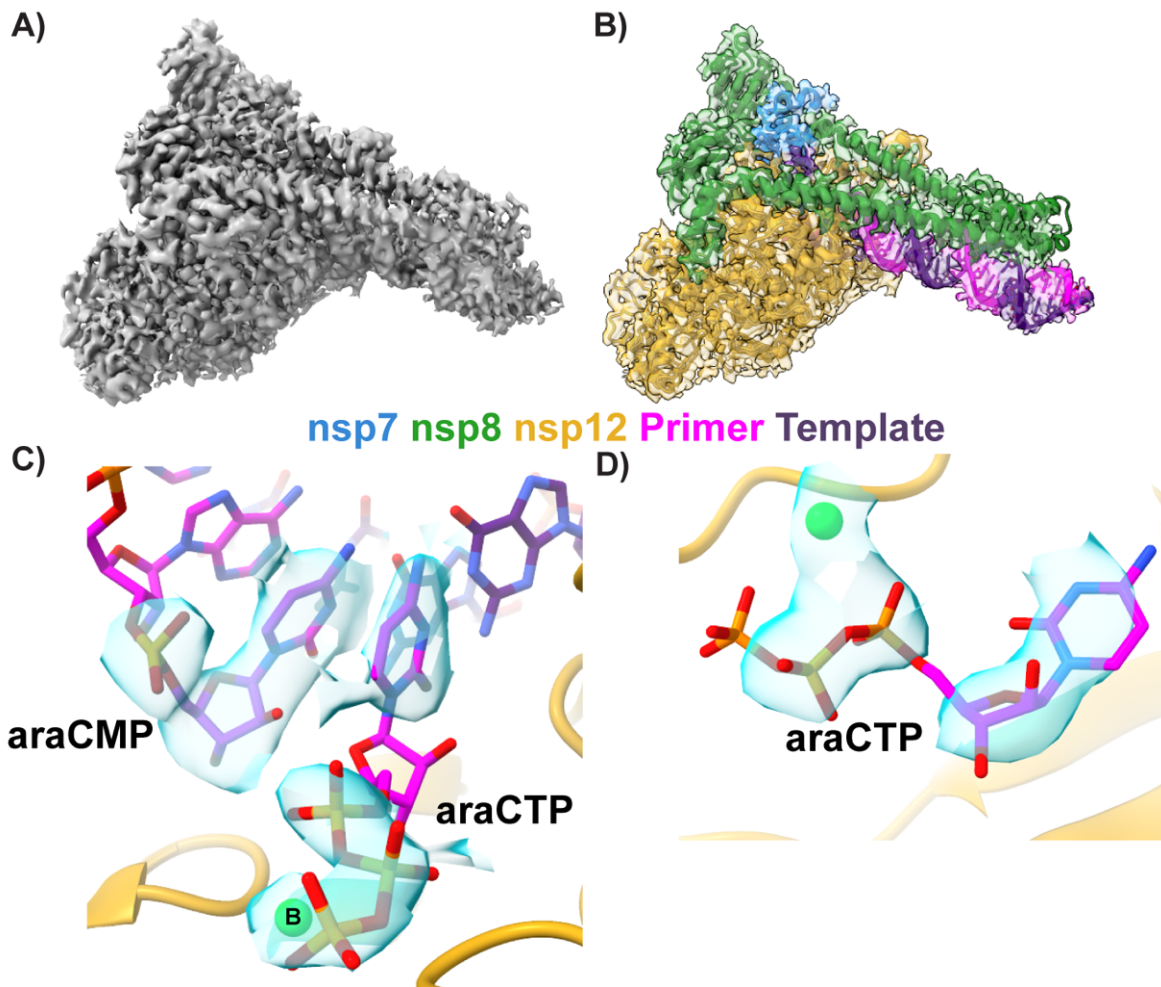


Figure 2, Structure of araCTP inhibited core-RTC: **A)** 3.3 Å cryoEM reconstruction of SARS-CoV-2 core-RTC. **B)** Atomic model of SARS-CoV-2 core-RTC docked into the cryoEM map colored by protein/RNA chain. **C and D)** araCMP and araCTP with surrounding electron density shown in light blue in the RdRP (**C**) and NiRAN (**D**) active sites. Mg²⁺ ions are displayed as bright green spheres.

Mechanism for araCTP inhibition of elongation

As noted above, there is an araCTP base-paired with the +1 base of template RNA in nsp12's RdRP active site. Much to our surprise the active site remained in an open conformation with the araCTP in this paired state (**Figure 3A**). This was unexpected as it is known that (+) ssRNA virus RdRPs, including nsp12, close their active site upon NTP binding (35,37). This active site closure is most notable with a shift in motif A towards the active site center. In our structure, motif A catalytic residue Asp618 is shifted 2.7 Å away from the polymerase active site when compared to a SARS-CoV-2 structure with CTP bound in a closed active site (**Figure 3A**) (37).

Work on picornavirus RdRPs demonstrated that the 2' and 3' hydroxyls of a bound NTP form a hydrogen bonding network with polymerase residues that is important for active site closure (35). The altered sugar pucker and positioning of the unincorporated araCTP result in an altered hydrogen bonding network compared to a structure with an unincorporated CTP (**Figure 3**) (37). The araCTP 2'OH loses hydrogen bonds to nsp12 residues Thr687 and Asn691 while Ser682 shifts (1.7 Å between α -carbons) to maintain a hydrogen bond with the 2'OH (**Figure 3B and C**). Residues Thr687 and Asn691 do not move during active site closure, so these observations are not affected by the different states of the two structures. Although interesting, these altered hydrogen bonds cannot be the main cause of araCTP inhibition as the first araCMP was incorporated efficiently. While the altered chirality and sugar pucker affects the 2'OH hydrogen bonds, the 3'OH remains near to nsp12-Asp623. While not directly bonding the araCTP 3'OH, nsp12-Asp623 is important for establishing the hydrogen bonding network with other nsp12 residues. The minimal change in distance between nsp12-Asp623 and the 3'OH indicates

that the unincorporated araCTP's C2' endo sugar pucker is not contributing to elongation inhibition (**Figure 3D and E**) (37).

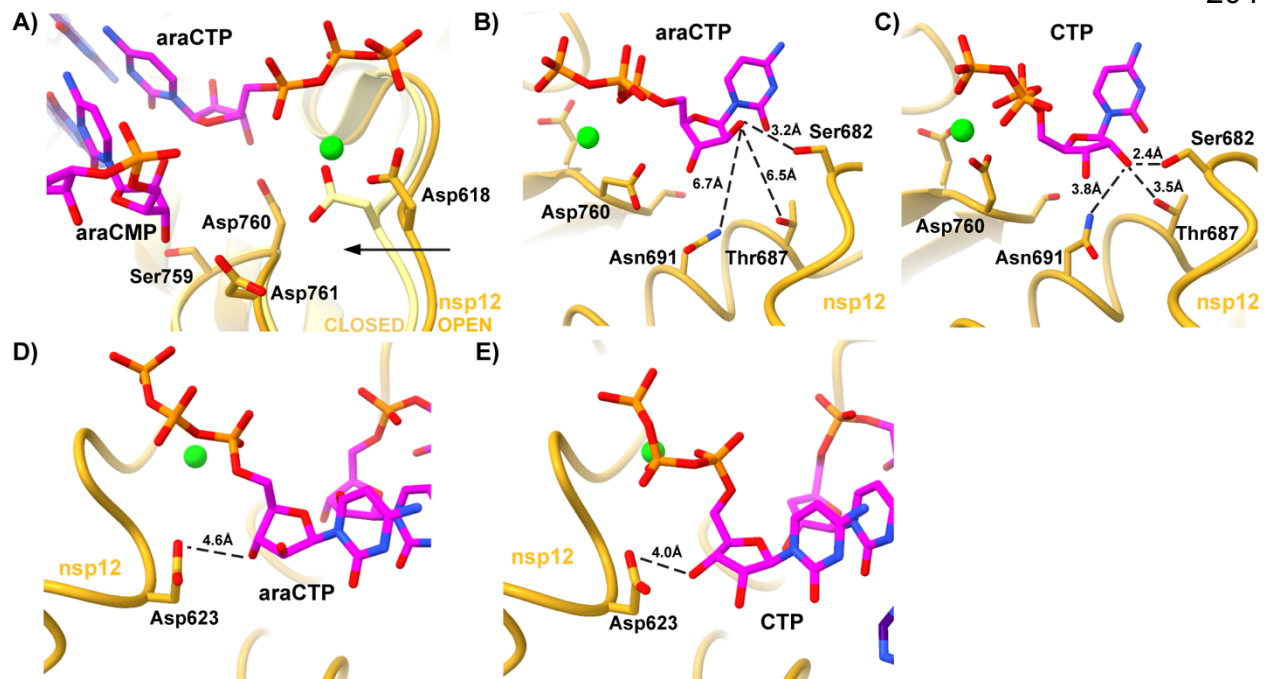


Figure 3, AraCTP in the RdRP active site: **A)** Superimposition of the SARS-CoV-2 RdRP active site in an open state (araCTP bound) and closed state (PDB: 7UOE). The shift in motif A to achieve the closed state is depicted by the black arrow and movement of nsp12-Asp618. **B and C)** Hydrogen bonding network of the 2'OH and 3'OH of araCTP (**B**) and CTP (**C**) (PDB: 7UOE). **D and E)** Distance between nsp12-Asp623, a residue important for hydrogen bonding network formation, and araCTP (**D**) or CTP (**E**). Distances between atoms are shown with black dashed lines and Mg^{2+} B ions are depicted as bright green spheres. PDB deposition 7UOE was used for figures of bound but unincorporated CTP.

We next analyzed the impact of the incorporated araCMP's altered chirality and C2' endo sugar pucker in the nascent RNA. The araCMP's 2'OH is oriented so that it cannot form a hydrogen bond with catalytic residue Ser759 during active site closure (**Figure 4A and B**). The arabinose conformation also orients the 3'OH such that it is pointing away from the incoming NTP α -phosphate (**Figure 4C and D**). Although the positioning of the araCMP vs cognate NMP 3'OH does not greatly shift, this altered orientation could contribute to araCMP inhibition.

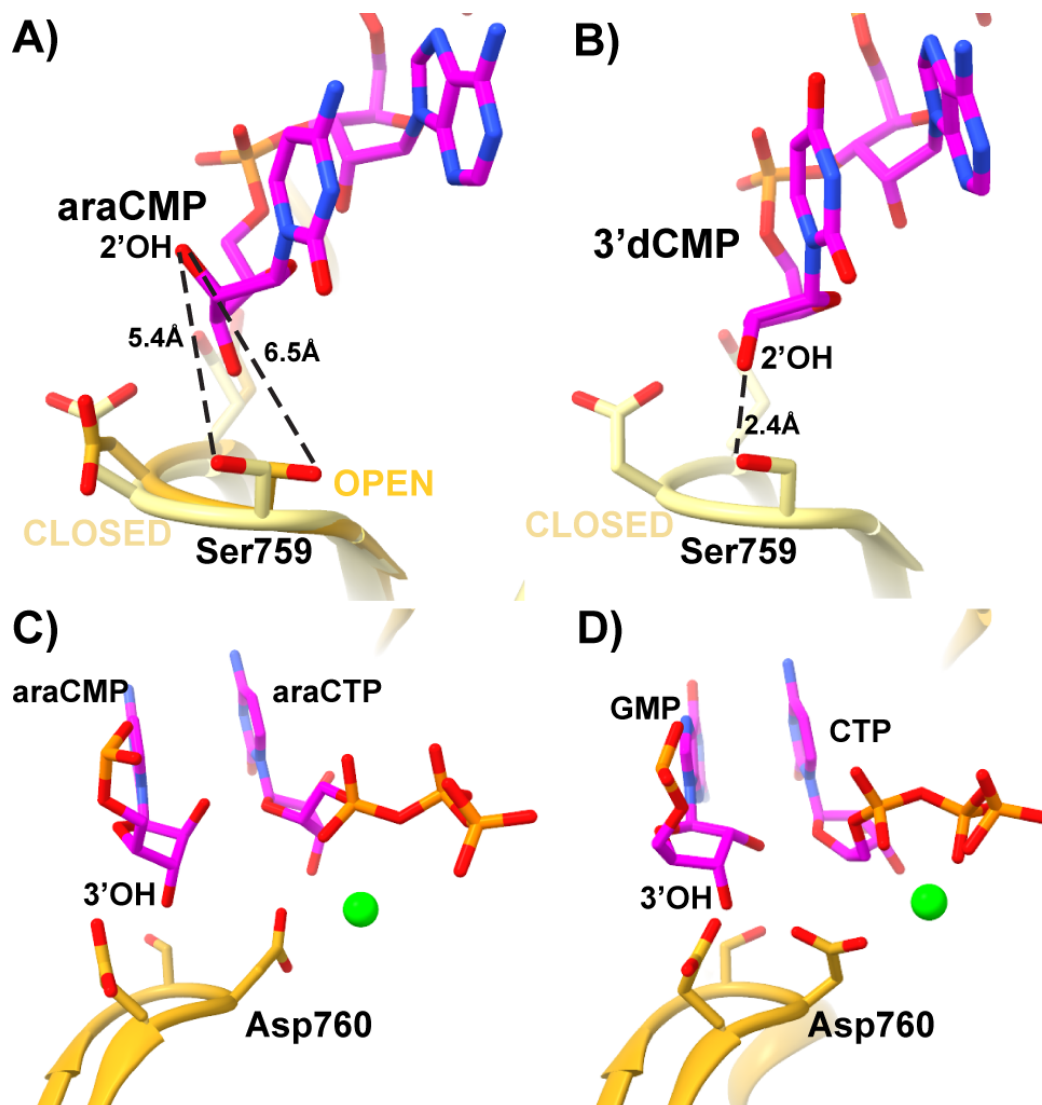


Figure 4, araCMP's altered sugar pucker inhibits RNA elongation: **A)** Superimposition of araCMP in open (gold) and closed (light yellow) active sites with the distance between Ser759 to the 2'OH indicated by dashed black lines. **B)** Structure of 3'dCMP in a closed active site (PDB: 7UOE) with the 2'OH forming a hydrogen bond (dashed black line) with nsp12-Ser759. **C and D)** The nucleophilic attack between the primer 3'OH and incoming nucleotide's α -phosphate could be inhibited by the altered orientation of the araCMP 3'OH imposed by the C2' endo sugar pucker (**C**) compared to an NMP C3' endo pucker (**D**). Figure (**D**) was produced by superimposing PDB models 6YYT and 7UOE to model a native primer (6YYT) bound to incoming CTP (7UOE). Mg²⁺ B ions are depicted as bright green spheres.

Discussion:

I have demonstrated that arabinose nucleotides are viable substrates for the SARS-CoV-2 polymerase that can potently block RNA elongation. Using single-particle cryoEM I solved the structure of a SARS-CoV-2 core-RTC whose elongation has been stalled after araCMP incorporation with another araCTP base-paired in an open polymerase active site. This structure has provided important insight into the possible mechanisms by which arabinose nucleotides inhibit the CoV replication machinery.

Trapping the SARS-CoV-2 RTC with an incorporated araCMP and unincorporated araCTP allowed us to demonstrate that arabinose nucleotides have several possible mechanisms by which they could block RNA extension. The unincorporated araCTP displayed an altered hydrogen bonding network that could be preventing motif A from shifting and active site closure. While we have shown that araCTP can be incorporated by the core RTC, its altered hydrogen bonds could still contribute to elongation stalling. The altered chirality of araCMP's 2' carbon orients the 2'OH such that it cannot form a hydrogen bond with active site residue nsp12-Ser759. The araCMP's C2' endo sugar pucker orients the primer 3'OH away from the araCTP α -phosphate and that could hinder nucleophilic attack or active site closure. With our current data, it is unclear which of these mechanisms, or combination thereof, are the key method(s) in which arabinose nucleotides inhibit the SARS-CoV-2 core-RTC. However, these structural observations demonstrate that arabinose nucleotides are strong elongation terminators of the CoV RdRP and deserve further evaluation as to their effectiveness against CoVs in tissue culture and animal models.

Arabinose nucleotides could also serve as a scaffold for the development of improved NAs that are able to target CoV polymerases more specifically and potently while reducing off-target effects. Fludarabine, an araATP with a fluorine at the position 2 carbon of the nitrogenous base, is a modified arabinose nucleotide that is used to treat different types of leukemia, in particular B-cell chronic lymphatic leukemias (65,66,68). Similar alterations to both the ribose, and nitrogenous base of araCTP and other arabinose nucleotides could be screened to identify novel and highly effective CoV antivirals.

As coronaviruses continue to crossover from animal reservoirs, the world will require antiviral drugs with diverse mechanisms of action by which we can halt virus spread in our efforts to prevent another pandemic. In addition to societal measures such as masking and social distancing, having an arsenal of antiviral drugs would allow us to quickly treat these new viral diseases, hopefully preventing future pandemics without the need for rapid vaccine design, testing, and implementation. Since we have now shown that arabinose nucleotides are substrates for the SARS-CoV-2 replication complex that can terminate RNA synthesis, we hope that these NAs will be pursued as potential CoV therapeutics.

References:

1. de Wit, E., van Doremalen, N., Falzarano, D. and Munster, V.J. (2016) SARS and MERS: recent insights into emerging coronaviruses.
2. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.

3. Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D. and Fouchier, R.A. (2012) Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. *N Engl J Med*, **367**, 1814-1820.
4. de Groot, R.J., Baker, S.C., Baric, R.S., Brown, C.S., Drosten, C., Enjuanes, L., Fouchier, R.A., Galiano, M., Gorbalenya, A.E., Memish, Z.A. *et al.* (2013) Middle East respiratory syndrome coronavirus (MERS-CoV): announcement of the Coronavirus Study Group. *J Virol*, **87**, 7790-7792.
5. Rabaan, A.A., Al-Ahmed, S.H., Sah, R., Alqumber, M.A., Haque, S., Patel, S.K., Pathak, M., Tiwari, R., Yattoo, M.I., Haq, A.U. *et al.* (2021) MERS-CoV: epidemiology, molecular dynamics, therapeutics, and future challenges. *Ann Clin Microbiol Antimicrob*, **20**, 8.
6. Wu, F., Zhao, S., Yu, B., Chen, Y.M., Wang, W., Song, Z.G., Hu, Y., Tao, Z.W., Tian, J.H., Pei, Y.Y. *et al.* (2020) A new coronavirus associated with human respiratory disease in China. *Nature*, **579**, 265-269.
7. (2024) WHO COVID-19 DASHBOARD.
8. Lyke, K.E., Atmar, R.L., Islas, C.D., Posavad, C.M., Szydlo, D., Paul Chourdury, R., Deming, M.E., Eaton, A., Jackson, L.A., Branche, A.R. *et al.* (2022) Rapid decline in vaccine-boosted neutralizing antibodies against SARS-CoV-2 Omicron variant. *Cell Rep Med*, **3**, 100679.
9. (2024) CDC. COVID-19 Treatment and Preventive Medication.
10. Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, **1282**, 1-23.
11. Yao, H., Song, Y., Chen, Y., Wu, N., Xu, J., Sun, C., Zhang, J., Weng, T., Zhang, Z., Wu, Z. *et al.* (2020) Molecular Architecture of the SARS-CoV-2 Virus. *Cell*, **183**, 730-738 e713.
12. Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.
13. Baranov, P.V., Henderson, C.M., Anderson, C.B., Gesteland, R.F., Atkins, J.F. and Howard, M.T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*, **332**, 498-510.
14. Plant, E.P. and Dinman, J.D. (2008) The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front Biosci*, **13**, 4873-4881.
15. Angelini, M.M., Akhlaghpour, M., Neuman, B.W. and Buchmeier, M.J. (2013) Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, **4**.

16. van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog*, **4**, e1000054.
17. Knoops, K., Kikkert, M., Worm, S.H., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M. and Snijder, E.J. (2008) SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. *PLoS Biol*, **6**, e226.
18. Malone, B., Urakova, N., Snijder, E.J. and Campbell, E.A. (2021) Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*.
19. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.
20. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
21. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
22. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2020) Cryo-EM Structure of an Extended SARS-CoV-2 Replication and Transcription Complex Reveals an Intermediate State in Cap Synthesis. *Cell*, **184**, 184-193.
23. 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. *et al.* (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.
24. Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L. and Denison, M.R. (2007) High fidelity of murine hepatitis virus replication is decreased in nsp14 exonuclease mutants. *J Virol*, **81**, 12135-12144.
25. Eckerle, L.D., Becker, M.M., Halpin, R.A., Li, K., Venter, E., Lu, X., Scherbakova, S., Graham, R.L., Baric, R.S., Stockwell, T.B. *et al.* (2010) Infidelity of SARS-CoV Nsp14-exonuclease mutant virus replication is revealed by complete genome sequencing. *PLoS Pathog*, **6**, e1000896.

26. Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. and 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.
27. Ma, Z., Pourfarjam, Y. and Kim, I.K. (2021) Reconstitution and functional characterization of SARS-CoV-2 proofreading complex.
28. Hillen, H.S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
29. Anderson, T.K., Hoferle, P.J., Chojnacki, K.J., Lee, K.W., Coon, J.J. and Kirchdoerfer, R.N. (2024) An alphacoronavirus polymerase structure reveals conserved replication factor functions. *Nucleic Acids Res.*
30. Černý, J., Černá Bolfíková, B., Valdés, J.J., Grubhoffer, L. and Růžek, D. (2014) Evolution of tertiary structure of viral RNA dependent polymerases. *PLoS One*, **9**, e96070.
31. Peersen, O.B. (2017) Picornaviral polymerase structure, function, and fidelity modulation. *Virus Res*, **234**, 4-20.
32. Appleby, T.C., Perry, J.K., Murakami, E., Barauskas, O., Feng, J., Cho, A., Fox, D., 3rd, Wetmore, D.R., McGrath, M.E., Ray, A.S. *et al.* (2015) Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase. *Science*, **347**.
33. Shu, B. and Gong, P. (2016) Structural basis of viral RNA-dependent RNA polymerase catalysis and translocation. *PNAS*, **113**.
34. Zamyatkin, D.F., Parra F Fau - Alonso, J.M.M., Alonso Jm Fau - Harki, D.A., Harki Da Fau - Peterson, B.R., Peterson Br Fau - Grochulski, P., Grochulski P Fau - Ng, K.K.S. and Ng, K.K. (2008) Structural insights into mechanisms of catalysis and inhibition in Norwalk virus polymerase. *Biological Chemistry*, **283**.
35. Gong, P. and Peersen, O.B. (2010) Structural basis for active site closure by the poliovirus RNA-dependent RNA polymerase. *PNAS*, **107**.
36. Temiakov, D., Patlan, V., Anikin, M., McAllister, W.T., Yokoyama, S. and Vassilyev, D.G. (2004) Structural basis for substrate selection by t7 RNA polymerase. *Cell*, **116**, 381-391.
37. Malone, B.F., Perry, J.K., Olinares, P.D.B., Lee, H.W., Chen, J., Appleby, T.C., Feng, J.Y., Bilello, J.P., Ng, H., Sotiris, J. *et al.* (2023) Structural basis for substrate selection by the SARS-CoV-2 replicase. *Nature*, 1-7.

38. Seley-Radtke, K.L. and Yates, M.K. (2018) The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Research*.
39. Jordan, P.C., Stevens, S.K. and Deval, J. (2018) Nucleosides for the treatment of respiratory RNA virus infections. *Antiviral Chemistry and Chemotherapy*.
40. Lau, J.Y., Tam Rc Fau - Liang, T.J., Liang Tj Fau - Hong, Z. and Hong, Z. (2002) Mechanism of action of ribavirin in the combination treatment of chronic HCV infection. *Hepatology*, **35**.
41. Witkowski Jt Fau - Robins, R.K., Robins Rk Fau - Sidwell, R.W., Sidwell Rw Fau - Simon, L.N. and Simon, L.N. (1972) Design, synthesis, and broad spectrum antiviral activity of 1- -D-ribofuranosyl-1,2,4-triazole-3-carboxamide and related nucleosides. *Journal of Medicinal Chemistry*, **15**.
42. Crotty, S., D, M., Jj, A., W, Z., Jy, L., Z, H., R, A. and Cameron, C.E. (2000) The broad-spectrum antiviral ribonucleoside ribavirin is an RNA virus mutagen. *Nature Medicine*.
43. Sidwell Rw Fau - Huffman, J.H., Huffman Jh Fau - Khare, G.P., Khare Gp Fau - Allen, L.B., Allen Lb Fau - Witkowski, J.T., Witkowski Jt Fau - Robins, R.K. and Robins, R.K. (1972) Broad-spectrum antiviral activity of Virazole: 1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide. *Science*, **177**.
44. Sofia, M.J. (2013) Nucleotide prodrugs for the treatment of HCV infection. *Advances in Pharmacology*, **67**.
45. 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. and 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.
46. Stockman, L.J., Bellamy R Fau - Garner, P. and Garner, P. (2006) SARS: systematic review of treatment effects. *PLOS Medicine*.
47. Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M. and Denison, M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog*, **9**, e1003565.
48. Gordon, C.J., Tchesnokov, E.P., Woolner, E., Perry, J.K., Feng, J.Y., Porter, D.P. and Gotte, M. (2020) Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J Biol Chem*, **295**, 6785-6797.

49. Sheahan, T.P., Sims, A.C., Graham, R.L., Menachery, V.D., Gralinski, L.E., Case, J.B., Leist, S.R., Pyrc, K., Feng, J.Y., Trantcheva, I. *et al.* (2017) Broad-spectrum antiviral GS-5734 inhibits both epidemic and zoonotic coronaviruses. *Sci Transl Med*, **9**.
50. Pruijssers, A.J., George, A.S., Schäfer, A., Leist, S.R., Gralinski, L.E., Dinnon, K.H., 3rd, Yount, B.L., Agostini, M.L., Stevens, L.J., Chappell, J.D. *et al.* (2020) Remdesivir Inhibits SARS-CoV-2 in Human Lung Cells and Chimeric SARS-CoV Expressing the SARS-CoV-2 RNA Polymerase in Mice. *Cell Reports*, **32**.
51. Beigel, J., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S. *et al.* (2020) Remdesivir for the Treatment of Covid-19 - Final Report. *The New England Journal of Medicine*, **383**.
52. Pan, H., Peto, R., Henao-Restrepo, A.M., Preziosi, M.P., Sathiyamoorthy, V., Abdool Karim, Q., Alejandria, M.M., Hernández García, C., Kieny, M.P., Malekzadeh, R. *et al.* (2021) Repurposed Antiviral Drugs for Covid-19 - Interim WHO Solidarity Trial Results. *The New England Journal Of Medicine*, **384**.
53. Yan, V.A.-O. and Muller, F.L. (2021) Why Remdesivir Failed: Preclinical Assumptions Overestimate the Clinical Efficacy of Remdesivir for COVID-19 and Ebola. *Antimicrobial Agents and Chemotherapy*, **65**.
54. Sheahan, T., Sims, A., Zhou, S., Graham, R., Pruijssers, A., Agostini, M.L., Leist, S., Schäfer, A., Dinnon, K.H.r., Stevens, L. *et al.* (2020) An orally bioavailable broad-spectrum antiviral inhibits SARS-CoV-2 in human airway epithelial cell cultures and multiple coronaviruses in mice. *Science Translational Medicine*, **12**.
55. Agostini, M.L., Pruijssers, A.J., Chappell, J.D., Gribble, J., Lu, X., Andres, E.L., Bluemling, G.R., Lockwood, M.A., Sheahan, T.P., Sims, A.C. *et al.* (2019) Small-Molecule Antiviral β -d-N (4)-Hydroxycytidine Inhibits a Proofreading-Intact Coronavirus with a High Genetic Barrier to Resistance. *J Virol*, **93**.
56. Jayk Bernal, A., Gomes da Silva, M.M., Musungaie, D.B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., Martín-Quirós, A., Caraco, Y., Williams-Diaz, A., Brown, M.L. *et al.* (2022) Molnupiravir for Oral Treatment of Covid-19 in Nonhospitalized Patients. *N Engl J Med*, **386**, 509-520.
57. Butler, C.C., Hobbs, F.D.R., Gbinigie, O.A., Rahman, N.M., Hayward, G., Richards, D.B., Dorward, J., Lowe, D.M., Standing, J.F., Breuer, J. *et al.* (2023) Molnupiravir plus usual care versus usual care alone as early treatment for adults with COVID-19 at increased risk of adverse outcomes (PANORAMIC): an open-label, platform-adaptive randomised controlled trial. *Lancet*, **401**, 281-293.

58. Kokic, G., Hillen, H.S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Hobartner, C. and Cramer, P. (2021) Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat Commun*, **12**, 279.
59. Kabinger, F., Stiller, C., Schmitzová, J., Dienemann, C., Kokic, G., Hillen, H.S., Höbartner, C. and Cramer, P. (2021) Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat Struct Mol Biol*, **28**, 740-746.
60. Sipkema, D., Franssen, M.C., Osinga, R., Tramper, J. and Wijffels, R.H. (2005) Marine sponges as pharmacy. *Mar Biotechnol (NY)*, **7**, 142-162.
61. Evich, M., Spring-Connell, A.M. and Germann, M.W. (2017) Impact of modified ribose sugars on nucleic acid conformation and function. *Heterocyclic Communications*, **23**, 155-165.
62. Seley-Radtke, K.L. and Yates, M.K. (2018) The evolution of nucleoside analogue antivirals: A review for chemists and non-chemists. Part 1: Early structural modifications to the nucleoside scaffold. *Antiviral Res*, **154**, 66-86.
63. Shepp, D.H., Dandliker, P.S. and Meyers, J.D. (1988) Current therapy of varicella zoster virus infection in immunocompromised patients. A comparison of acyclovir and vidarabine. *Am J Med*, **85**, 96-98.
64. Reese, N.D. and Schiller, G.J. (2013) High-dose cytarabine (HD araC) in the treatment of leukemias: a review. *Curr Hematol Malig Rep*, **8**, 141-148.
65. Shah, N.N., Lee, D.W., Yates, B., Yuan, C.M., Shalabi, H., Martin, S., Wolters, P.L., Steinberg, S.M., Baker, E.H., Delbrook, C.P. *et al.* (2021) Long-Term Follow-Up of CD19-CAR T-Cell Therapy in Children and Young Adults With B-ALL. *J Clin Oncol*, **39**, 1650-1659.
66. Dekker, L., Calkoen, F.G., Jiang, Y., Blok, H., Veldkamp, S.R., De Koning, C., Spoon, M., Admiraal, R., Hoogerbrugge, P., Vormoor, B. *et al.* (2022) Fludarabine exposure predicts outcome after CD19 CAR T-cell therapy in children and young adults with acute leukemia. *Blood Adv*, **6**, 1969-1976.
67. Adkins, J.C., Peters, D.H. and Markham, A. (1997) Fludarabine. An update of its pharmacology and use in the treatment of haematological malignancies. *Drugs*, **53**, 1005-1037.
68. Plosker, G.L. and Figgitt, D.P. (2003) Oral fludarabine. *Drugs*, **63**, 2317-2323.
69. Punjani, A., Rubinstein, J.L., Fleet, D.J. and Brubaker, M.A. (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods*, **14**, 290-296.

70. Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J.A.-O. and Ferrin, T.A.-O. (2021) UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Science*, **30**, 70-82.
71. Emsley, P., Lohkamp, B., Scott, W.G. and Cowtan, K. (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr*, **66**, 486-501.
72. Croll, T.I. (2018) ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol*, **74**, 519-530.
73. Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A. and Adams, P.D. (2018) Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol*, **74**, 531-544.
74. Steitz, T.A. (1998) A mechanism for all polymerases. *Nature*, **391**, 231-232.

Supplemental Data:

EMDB	44654 (held for publication)
PDB	9BLF (held for publication)
Microscope	Talos Arctica
Voltage (kV)	200
Detector	K3 direct electron detector (Gatan)
Dose Rate (e ⁻ /pixel/sec)	14.2
Exposure Time (sec)	4.82
Electron Exposure (e ⁻ /Å ²)	60
Frames (no.)	60
Defocus Values	-0.5, -1.0, -1.5, -2.0
Data Collection Mode	EFTEM, Counting, CDS
Nominal Magnification	79,000
Pixel Size (Å)	1.064
Symmetry Imposed	C1
Movies Collected (no.)	4,230
Initial Particle Images (no.)	1,321,870
Final Particle Images (no.)	158,606
Map Resolution (Å) – GSFSC	3.3
Initial Models Used (PDB ID)	7UOE
Non-hydrogen Atoms	12,293
Protein Residues	1,375
Nucleic Acid Residues	63
Other Atoms	2 (Zn ²⁺), 2 (Mg ²⁺)
Ligands	2 HF4 (araCTP), 1 CAR (araCMP)
R.M.S. Deviations	
Bond Lengths (Å)	0.006
Bond angles (°)	0.498
MolProbity Score	1.50
Clashscore	3.00
Ramachandran Plot	
Favored (%)	97
Allowed (%)	3
Disallowed (%)	0

Table S1, cryo-EM data collection and refinement: Information provided is for the cryoEM data collection, and processing that produced the electron density map of the araCTP inhibited SARS-CoV-2 polymerase complex. Validation statistics for the polymerase complex coordinate model built into the reconstruction are provided.

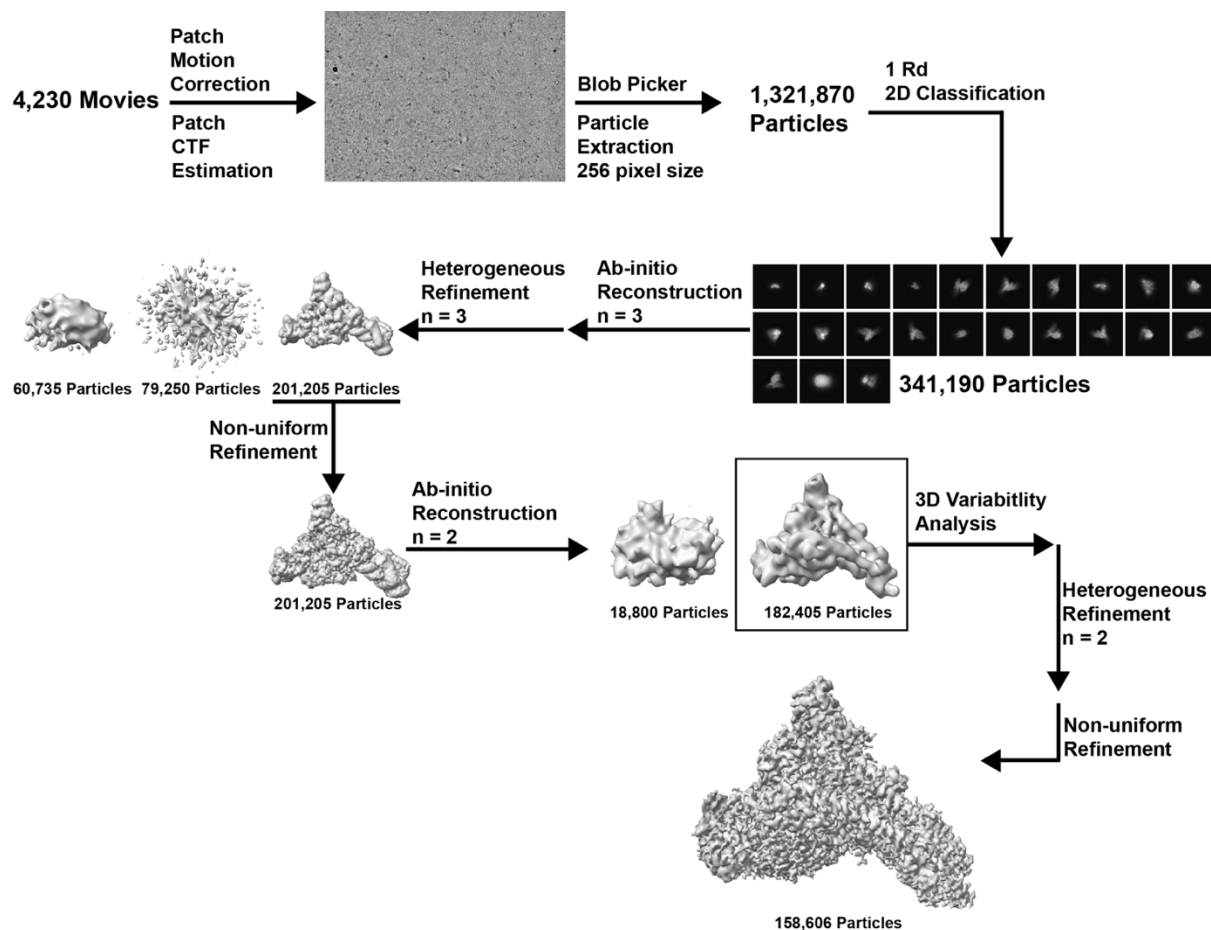


Figure S1, CryoEM Data Processing Pipeline: Workflow for the processing of cryoEM data that resulted in our final reconstruction. Data was processed using cryoSPARC. The final resolution of our reconstruction is 3.3 Å.

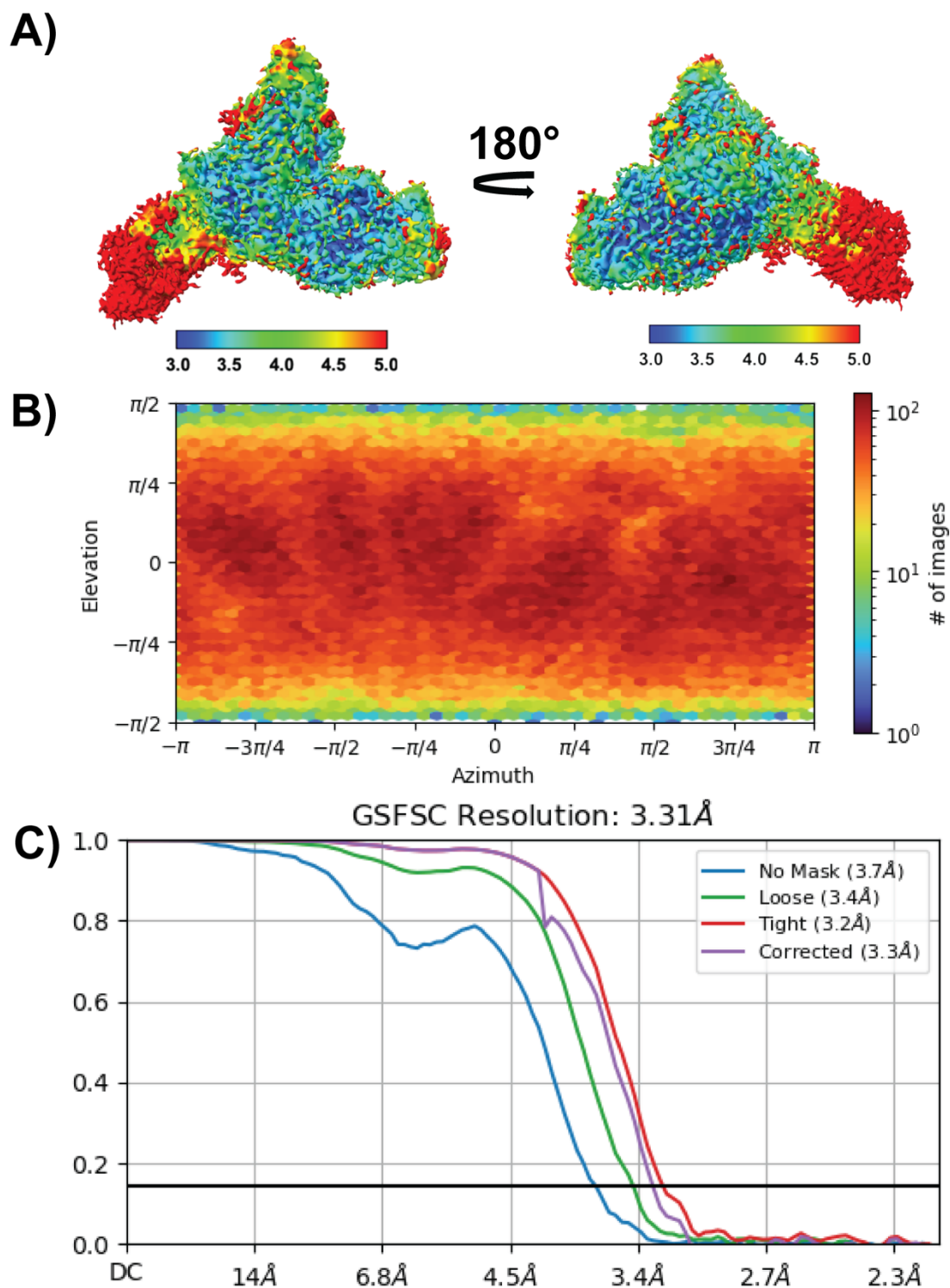


Figure S2, CryoEM Validation: **A)** CryoEM reconstruction colored by local resolution. **B)** Projection orientation distribution of final reconstruction. “# of Images” is the number of projections at each orientation. **C)** Gold-standard Fourier shell correlation (FSC), horizontal black line across plot indicates 0.143 cutoff.

Chapter 6: Discussion and Future Directions

Authors: Thomas K. Anderson¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A. wrote this chapter with edits by R.N.K.

Summary:

For a virus to successfully replicate within a cell it requires both viral and host encoded machinery. Many DNA viruses are able to coopt host DNA replication and transcription machinery, benefitting from the fact that they share a genetic makeup of DNA with their infected host. RNA viruses are not so fortunate and must encode their own RNA dependent polymerases to produce viral RNA products during an infection. Viral RNA-dependent RNA polymerases are virus specific replication factors that can be specifically targeted to block virus infections. Beyond the RNA polymerase, RNA viruses employ countless different mechanisms to efficiently replicate within a cell. This includes the different enzymes and replication factors these viruses have evolved and encode in their genomes. RNA virus genomes are constrained in size, so the enzymes and replication factors present serve critical roles during virus infection.

Among RNA viruses, coronaviruses (CoV) have some of the largest known genomes (1). With these large genomes comes complex mechanisms of viral RNA replication. CoVs replicate their genomes using a replication-transcription complex (RTC) that is believed to include over half a dozen viral enzymatic functions and several other viral and host replication factors (2). How these viral enzymes and replication factors interact to mediate the production and modification of viral RNA has been the focus of much work. As we continue to learn about the dynamics and interactions of the CoV RTC, it becomes readily apparent that there is still much to learn about the complicated mechanisms of CoV replication.

In the second chapter of my thesis, we sought to address the lack of structural and biochemical knowledge on non-*BetaCoV* RTCs. In doing so, we solved the first two

structures of CoV RTCs outside of the *BetaCoV* genera. Our two structures of the *AlphaCoV* core-RTC were solved with the RTC bound to RNA, in the absence or presence of a viral replication factor that was thought to be critical for RNA binding and synthesis. Further characterization of this observation led to the conclusion that this cofactor is not required for RNA synthesis, and likely has other roles in replication.

As discussed above we are interested in expanding our knowledge on CoV RTCs beyond the *BetaCoVs*. Building upon the work in **Chapter 2**, in the third chapter we solved the first structure of a *GammaCoV* core-RTC. We observed several structural differences between the *GammaCoV* RTC and those of *Alpha- and BetaCoVs*. Most notably, we saw altered interactions between the polymerase and a replication factor across all three genera RTCs. Further analysis of these altered interactions identified the potential of genus-specific mechanisms for RTC assembly.

Moving beyond the core-RTC, in the fourth chapter of my thesis I aimed to elucidate the unique mechanism of CoV proofreading. While proofreading is essential for increased fidelity during replication, the mechanism by which the viral proofreading enzyme, nsp14, mediates proofreading is unknown. We determined that nsp14 interacts with the viral RTC, and that this interaction is dependent on the RTC being in a backtracked state. This work provides critical preliminary insight into the mechanism of CoV proofreading.

In the fifth chapter we explored the potential of arabinose nucleotides in their ability to block CoV RNA synthesis. Arabinose nucleotides are historically significant nucleoside analogues that had not yet been explored in their potential to treat CoV induced diseases. Using biochemical assays and structural biology we identified that

arabinose nucleotides potentially block SARS-CoV-2 elongation. This work provides a cornerstone from which arabinose nucleotide inhibitors could be developed for CoV disease treatment.

In summation, the work presented here expands the field of CoV RTC biology to understudied branches of the CoV subfamily, and in so doing, demonstrates that critical insight into CoV replication can be gained by exploring understudied viruses. Further, we have begun to elucidate the unique mechanism by which CoVs proofread during replication.

1.0 The coronavirus core-RTC:

As mentioned above, the CoV RTC is believed to be made up several viral enzymes and cofactors. The central factor of the RTC is the viral protein nsp12, which contains both the viral RNA-dependent RNA polymerase and a nucleotidyltransferase domain (3,4). Nsp12 assembles with viral cofactors nsp7 and nsp8 into the core-RTC, the minimum complex needed for processive RNA synthesis *in vitro* (3). Preliminary structures of the CoV core-RTC showed that it was composed of one nsp12, one nsp7, and two nsp8s (5). Each nsp8 has a long helical extension that binds dsRNA out of the polymerase active site (6). Further structures, including our own, have shown that the core-RTC can remain bound to RNA in the absence of one of the nsp8s (Chapter 2 and (7,8)). Since RdRPs are virus specific enzymes, they are often the target of nucleoside analogues. Nsp12 and the core-RTC are no exception and have been the focus of antiviral development. Currently there are only two nucleoside analogues approved for the treatment of CoV induced diseases in humans.

1.1 Core RTCs from different genera:

CoV nsps, in particular those involved in genome replication (i.e. nsp7, and nsp12), are highly conserved (~50% amino acid identity) across the CoV subfamily. This high sequence similarity indicates that the different CoV genera replicate using similar, if not identical mechanisms. Structures of *BetaCoV* RTCs have illuminated several important aspects of CoV replication. Expanding structural knowledge beyond *BetaCoV* polymerases could confirm the hypothesis that CoV genera replicate using similar mechanisms, while also potentially identifying previously unknown aspects of replication. To expand our knowledge into other CoV genera, we solved the first two structures of *AlphaCoV* core-RTCs (Chapter 2) and the first structure of a *GammaCoV* core-RTC (Chapter 3).

In studying the core-RTC of the *AlphaCoV* PEDV we solved two structures: one with nsp8_T bound and one without (Chapter 2). Interestingly, both complexes remained bound to RNA in the presence or absence of this viral cofactor. This was initially surprising as these nsp8s have been shown to be critical for RTC binding to, and extension of RNA substrates (3). The work that established the importance of nsp8 in binding and extending RNA lacked the structural knowledge that two nsp8s interact with the RTC, binding dsRNA leaving the active site (Chapters 2 and 3, and (6)). Using a previously established cofactor fusion protein (nsp8L7) we delineated the contributions of each nsp8 in RTC polymerase activity, revealing that nsp8_T's RNA binding domain is not required for *in vitro* RNA synthesis for PEDV (Chapter 2 and (9)). Further, we and others demonstrated that this is a conserved feature with the SARS-CoV-2 RTC (Chapter 2 and (9)).

Our structure of the *GammaCoV* IBV's core-RTC revealed several differences in the RTC structure when compared to those of PEDV and SARS-CoV-2 (Chapter 3). Among these differences were multiple loops that were either extended or shortened. These differences occurred primarily at sites distal to known nsp binding sites. The most notable difference was a drastic conformational change in a loop in the nsp12 interface domain that interacts with the nsp8_F head domain in PEDV and IBV, but not for SARS-CoV-2. We went on to demonstrate that interaction between this loop and nsp8_F is critical for RTC assembly for polymerases of *Alpha*- and *Gamma*-, but not *BetaCoV*.

By solving the structure of core-RTCs from several different genera we saw that the overall structure of CoV RTCs is highly conserved, and in particular the nsp12 active sites have high structural conservation. This conservation indicates that drugs targeting the CoV RdRP or NiRAN could be broadly acting. Our work on PEDV and IBV revealed important insight into nsp12-nsp8 interactions. In both Chapter 2 and Chapter 3 we demonstrate that nsp8_F is the critical cofactor in RTC assembly, RNA binding, and polymerase activity. Others have shown that nsp8_F binding to nsp12 is the rate limiting step in RTC activation (9). Compared to nsp8_F, our work and others has begun to disprove the theory that nsp8_T is required for RNA binding and synthesis (Chapter 2 and (9)). Further *in vitro* characterization of these two nsp8s is required to delineate their roles and necessity in RTC function. The mutant and fusion nsps described in our work (Chapter 1) could be tested on long dsRNA substrates to probe the nsp8s function in processive RNA synthesis. Prior work in the field has established processive elongation assays using single molecule techniques. Further, the effects of nsp8 domain interaction-disrupting mutations (Chapter 2) could be tested on virus replication using reverse genetics. These

experiments could examine their broader impact on other aspects of CoV replication. In addition, our work on the IBV RTC revealed that there are structural differences between genera that have functional consequences (Chapter 3). To further examine the differences between genera more structures of more diverse CoV RTCs could be solved and explored.

1.2 Interpreting structures of stalled RTCs:

Structural biology has proven to be an important tool in studying CoV RTCs. This includes the thesis work presented in Chapters 2, 3, and 5. While structures of CoV RTCs provide critical insight, it is important to remember that to solve these structures, samples must be highly uniform and stable complexes. This often involves solving the structures of RTCs in stalled states. Interpretation of these structures should take into consideration the fact that they are stalled, and follow-up experiments should be done to test structural observations. A good example of this is that it was initially predicted that both nsp8s' N-terminal extensions were required for RNA binding and elongation based off preliminary biochemical and structural observations. But further structural and biochemical experimentation done by us and others has demonstrated that only one nsp8 N-terminal extension is required for RNA synthesis *in vitro* (Chapter 2, (9)).

When thinking about the structures of CoV RTCs with two nsp8 extensions binding dsRNA leaving the active site, a question arises, what state is this RTC in? This RTC could either be a stalled elongation complex and/or a stalled state that has another function such as initiating backtracking or proofreading. Based on our results that nsp8 τ 's RNA binding domain is not required for RNA synthesis, I predict that these structures are paused elongation complexes, rather than actively elongating (**Figure 1**). Whether

pausing is promoted by, or permits nsp8_T binding to dsRNA remains unknown, but this association allows other viral processes such as backtracking (mediated by nsp13 bound to nsp8_T) to occur (10). In support of nsp8_T's role in pausing, preliminary experiments have shown that mutations designed to disrupt nsp8_T's N-terminal association with the RTC promote a decrease in pre-misincorporation pausing (Appendix 1). To further test the regulation of RTC stalling mediated by nsp8_T, *in vitro* assays that force pausing (i.e. mismatch introduction or RNA secondary structure) should be more thoroughly executed. Further, to delineate the two nsp8s' contributions to pausing, the ns8L7 fusion protein could be used, screening different mutations that block, hinder, or remove nsp8_T's RNA binding ability. These experiments could be performed on both long and short RNA substrates to access the role of pausing on processivity as well.

1.3, The CoV RTC in the context of infection:

Until now I have focused my discussion on the core-RTC (nsp7+nsp8+nsp12), but CoV replication involves many more enzymes and cofactors beyond those of the core-RTC. This includes the viral helicase nsp13, for which two protomers bind the RTC via interactions with each nsp8 (11). Nsp8_T, which we have discussed as not being needed for RNA synthesis, scaffolds the nsp13 that can mediate CoV backtracking (10). We predict that nsp8_T regulates CoV backtracking (**Figure 1**). To test nsp8_T's potential role in backtracking, site directed mutants on nsp8 that block the backtracking nsp13 binding to the RTC could be screened with *in vitro* backtracking assays and using live virus and reverse genetics. Sequencing of the RNA produced by mutant viral strains could provide better context for the role of CoV backtracking, including its potential part in CoV proofreading as discussed in Chapter 4.

Other viral enzymes and their interactions with the RTC are relatively uncharacterized compared to nsp13. Our work in isolating the proofreading-RTC (Chapter 4) is beginning to piece together the interactions between the RTC and the exonuclease complex. It is poorly understood if and how the viral endonuclease (nsp15) and 2'O methyltransferase (nsp16) interact with the RTC. To examine the potential interactions of the core-RTC with these various nsps, permutations of nsps (i.e. proofreading-RTC +/- nsp13) could be tested for complex formation using RNA substrates that depict the different stages of RNA replication (i.e. backtracked, stalled...). Similar work was presented in Chapter 4 to identify the proofreading-RTC.

An aspect of the RTC that is not discussed in detail in this thesis is the RTCs seclusion in membrane bound viral replication vesicles during infection (12). These vesicles have been shown to have large viral protein pores that span the membranes of the vesicles, allowing the egress of viral RNA products for translation and packaging (13). Within the context of these vesicles, it is unknown how the RTC functions, including the amount of RTCs present and if the RTC(s) interact with the pore complex. To better characterize RTC dynamics in replication vesicles, cryo-electron tomography could be used on infected cells, or isolated replication vesicles, to observe the RTC in a more native state.

2.0 Coronavirus proofreading and nucleoside analogues:

CoVs and other closely related viruses in the order *Nidovirales* are the only known RNA viruses that proofread during replication. For CoVs, proofreading has been shown to be critical for increasing fidelity during replication (14). The development of nucleoside analogues against CoVs has proven difficult due to CoVs ability to proofread. As

discussed in Chapters 4 and 5, to be an effective nucleoside analogue the antiviral must both be efficiently incorporated by the RTC and evade excision by nsp14.

2.1 Interaction between the RTC and exonuclease complex:

Prior to the work presented in Chapter 4, there were few reports of interactions between the RTC and nsp14, and these reports were contradictory to one another and often used fusion proteins that potentially promoted nsp14 binding to the RTC when it otherwise wouldn't (3,15). To address these contradictory results, we identified that the core-RTC and exonuclease complex (nsp10+nsp14) can interact, and that the interaction is dependent on RTC binding of backtracked RNA (Chapter 4). We then demonstrated that there is a window of backtracked RNA that permits interaction of the two complexes and assembly of the proofreading-RTC (Chapter 4). While we eliminated the NTP channel as the potential binding site of the exonuclease complex on the RTC, we did not go on to identify the location of the interaction (Chapter 4). Among the possible sites for exonuclease complex binding is at or near the nsp7-nsp8 heterodimer. This site is an attractive option as it is still close enough that if nsp10+nsp14 bound there it could reach 8-14 nts of backtracked RNA. In addition, it has been previously shown that nsp8 can bind nsp14 or the complex of nsp10+nsp14 (16,17). To determine the site of interaction between the RTC and exonuclease complex a more thorough mutagenic screen attempting to disrupt complex assembly could be performed. Further, a structure of the two complexes interacting on a backtracked RNA substrate would provide critical details about the interaction.

The dependence on a backtracked RNA substrate leads us to predict that nsp13 mediated backtracking is involved in proofreading (**Figure 1**). But it is important to state

that this has not been proven yet. Interestingly, a recent publication tested the viability of a CoV that lacked a protease cleavage site between nsp13 and nsp14, resulting in nsp13 and nsp14 only being present as a fusion protein during infection (18). This mutation resulted in a viable virus with a mutation rate in-between that of wildtype and proofreading deficient viruses. These results indicate that nsp13 and nsp14 could associate with the RTC at the same time, potentially during proofreading that would require nsp13 mediated backtracking. To test if backtracking is involved in proofreading, an *in vitro* proofreading assay could be developed, and mutations blocking nsp13 binding to the RTC could be tested for their impact on proofreading efficiency. In addition, backtracking defective viral strains could be developed using reverse genetics, and virus mutation rates could be examined in these strains.

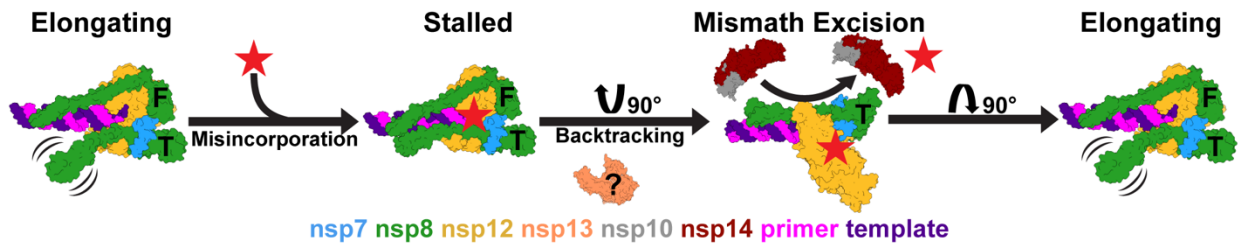


Figure 1, Hypothetical model of CoV proofreading: A misincorporation (red star) is hypothesized to initiate stalling of the CoV core-RTC. Stalling is predicted to require or promote nsp8_T's N-terminal extension binding to dsRNA and allow backtracking by the RTC. While nsp13 is known to mediate CoV backtracking, whether it is involved in proofreading is currently not known. Backtracking by the CoV RTC results in the 3' end of nascent RNA being extruded out the NTP channel, allowing the exonuclease complex to bind the RTC, and excise the mismatch. After mismatch excision, the exonuclease complex disassociates from the RTC allowing RNA to re-enter the active site and elongation to continue.

2.2 Evolution of CoV proofreading:

The CoV proofreading enzyme, nsp14, is a 3'-5' exoribonuclease in the DE(E/D)D superfamily of exonucleases (19). This is the same superfamily of exonucleases as those involved in DNA proofreading in eukaryotes and prokaryotes. Other virus families with genomes ≥ 20 kb in the *Nidovirales* order encode a homologous exonuclease to nsp14. Virus families in *Nidovirales* with genomes < 20 kb, such as *Arteriviridae*, lack a homologue for nsp14 and do not proofread during replication. Given that there are viruses within *Nidovirales* with genomes < 20 kb that lack a proofreading enzyme it is reasonable to hypothesize that the ability to proofread was acquired within the *Nidovirales* lineage. Within *Nidovirales*, the viruses with the smallest genomes (20 kb) that proofread during replication are the *Mesoniviridae* family. Perhaps an ancestral nidovirus acquired the ability to proofread allowing their genomes to increase to, and past, 20 kb. Another possibility is that proofreading was present in the *Nidovirales* last common ancestor, and that an ancestor to arteriviruses lost their ability to proofread, causing a shift towards a smaller genome size for that family. While both theories are interesting, it is currently unknown which of them (or others) describes the evolution of *Nidovirales* proofreading. Sequencing of a broad range RNA viruses has already begun to identify more viruses within *Nidovirales*, and these continued efforts could help piece together the evolution of *Nidovirales* proofreading.

Further complicating our understanding of *Nidovirales* proofreading is the lack of an identified ancestral origin for nsp14. Belonging to the same superfamily as DNA proofreading enzymes hints that nsp14's ancestor could be a DNA proofreading enzyme, the gene for which was acquired during an infection. But an ancestral link between a

known DNA proofreading enzyme and nsp14 has yet to be shown. To begin to better understand the ancestral origins of nsp14 I believe that more work should be done to characterize the exonuclease activity of non-CoV proofreading Nidoviruses (i.e. *Mesoniviridae*). This work could identify structural and biochemical differences that define an evolutionary development between the *Nidovirales* exonucleases.

3.0 Concluding remarks:

CoVs are RNA viruses with large genomes that encode numerous enzymes and replication factors that interact to synthesize, proofread, and modify viral RNA. Because of this, CoV replication is a highly complicated process, the mechanisms for which we are just beginning to understand. In this thesis, we contributed to the growing field of CoV replication biology by expanding our structural and biochemical understanding to the understudied *Alpha-* and *GammaCoV polymerases*. Further, we have begun to elucidate the previously poorly characterized mechanism of CoV proofreading. As CoVs continue to pose a threat to societal health, having a more complete understanding of their replication processes will help us treat current and future CoV induced diseases.

References:

1. Gorbalenya, A.E., Enjuanes, L., Ziebuhr, J. and Snijder, E.J. (2006) Nidovirales: evolving the largest RNA virus genome. *Virus Res*, **117**, 17-37.
2. van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog*, **4**, e1000054.
3. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.

4. Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res*, **96**, 59-126.
5. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
6. Hillen, H.S., Kobic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.
7. Anderson, T.K., Hoferle, P.J., Chojnacki, K.J., Lee, K.W., Coon, J.J. and Kirchdoerfer, R.N. (2024) An alphacoronavirus polymerase structure reveals conserved replication factor functions. *Nucleic Acids Res*.
8. Bravo, J.P.K., Dangerfield, T.L., Taylor, D.W. and Johnson, K.A. (2021) Remdesivir is a delayed translocation inhibitor of SARS-CoV-2 replication. *Molecular Cell*, **81**, 1548-1552.
9. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.
10. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, **118**.
11. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T. *et al.* (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
12. Angelini, M.M., Akhlaghpour, M., Neuman, B.W. and Buchmeier, M.J. (2013) Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. *mBio*, **4**.
13. Wolff, G., Limpens, R., Zevenhoven-Dobbe, J.C., Laugks, U., Zheng, S., de Jong, A.W.M., Koning, R.I., Agard, D.A., Grünewald, K., Koster, A.J. *et al.* (2020) A molecular pore spans the double membrane of the coronavirus replication organelle. *Science*, **369**, 1395-1398.
14. Eckerle, L.D., Lu, X., Sperry, S.M., Choi, L. and Denison, M.R. (2007) High fidelity of murine hepatitis virus replication is decreased in nsp14 exoribonuclease mutants. *J Virol*, **81**, 12135-12144.
15. Yan, L., Yang, Y., Li, M., Zhang, Y., Zheng, L., Ge, J., Huang, Y.C., Liu, Z., Wang, T., Gao, S. *et al.* (2021) Coupling of N7-methyltransferase and 3'-5'

- exoribonuclease with SARS-CoV-2 polymerase reveals mechanisms for capping and proofreading. *Cell*, **184**, 3474-3485.e3411.
16. Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B. and Yang, Y. (2021) Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science*, **373**, 1142-1146.
 17. von Brunn, A., Teepe, C., Simpson, J.C., Pepperkok, R., Friedel, C.C., Zimmer, R., Roberts, R., Baric, R. and Haas, J. (2007) Analysis of intraviral protein-protein interactions of the SARS coronavirus ORFeome. *PLoS One*, **2**, e459.
 18. Anderson-Daniels, J.A.-O., Gribble, J. and Denison, M. Proteolytic Processing of the Coronavirus Replicase Nonstructural Protein 14 Exonuclease Is Not Required for Virus Replication but Alters RNA Synthesis and Viral Fitness.
 19. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.

Appendix 1: Isolation of a stalled pre-misincorporation SARS-CoV-2 RTC

Authors: Thomas K. Anderson¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A. generated all data presented in this appendix.

T.K.A. wrote this appendix with editing by R.N.K..

Introduction:

Coronaviruses (CoVs) are (+) single-stranded RNA viruses with large, ~30 kb genomes (1). The 5' two-thirds of the CoV genome encodes for 16 viral non-structural proteins, or nsps (2). Several of these nsps interact, forming the viral replication-transcription complex (RTC) that synthesizes and modifies viral RNA products during an infection (3). The viral RNA-dependent RNA polymerase (RdRP) nsp12 is the central component of the RTC (4). Nsp12, with replication factors nsp7 and nsp8, form the minimum complex needed for processive RNA synthesis *in vitro* (5,6). Because of this, the complex composed of all three nsps is referred to as the core-RTC. In addition to nsp12, CoVs encode several other enzymatic nsps that interact, or are predicted to interact, with the core RTC (7). Among these is the viral RNA helicase nsp13 (8).

Structural studies of several CoVs revealed a surprising core-RTC stoichiometry of one nsp12 bound to one nsp8 and one nsp7+nsp8 heterodimer (9,10). The nsp8s are referred to as nsp8_T (nsp7 heterodimer) and nsp8_F depending on if they interact with the polymerase thumb or fingers domain, respectively. Structures of CoV core-RTCs bound to RNA duplexes have shown that long N-terminal extensions of the two nsp8s bind upstream dsRNA exiting the RdRP active site (10,11). This interaction has been proposed to promote core-RTC RNA binding and processive elongation (10). The CoV helicase, nsp13, is a superfamily 1B helicase that is important for virus replication (12-14). Structures of nsp13 bound to the core-RTC have shown that two helicase protomers are scaffolded to the RTC, each by one of the nsp8s (15,16). The nsp13 scaffolded by nsp8_T was shown to bind the 5' end of template RNA that would be entering the RdRP active site to be copied during RNA elongation (15,16). This interaction is interesting as the direction of the helicase translocation on template RNA (5'-3') is opposite that of the polymerase, nsp12 (3'-5'). Further structural and biochemical characterization determined that this nsp13 can mediate CoV RTC backtracking, which results in the 3' end of primer/nascent RNA being extruded out of nsp12's NTP channel (17). At present, the role and regulation of nsp13-mediated RTC backtracking is unknown, but it has been predicted that backtracking could be involved in CoV discontinuous strand synthesis or proofreading (17). In addition, the purpose of the second nsp13 protomer, scaffolded to the RTC by nsp8_F, is unknown.

To solve the structure of the backtracked RTC, an RNA duplex with 5 consecutive mismatches was used (17). Although unlikely to occur naturally, the use of several consecutive mismatches on an RNA duplex is commonly used to stably trap polymerases in a backtracked state, allowing structural elucidation (17,18). The biochemical validation of nsp13 mediated backtracking was done using a pre-assembled and stalled RTC (17). To this stalled RTC, nsp13 and ATP were added allowing nsp13 to bind the RTC and drive backtracking of the previously non-elongating complex. Though providing new insights into nsp13, trapping the RTC in a backtracked state for structure solving or inducing backtracking on an already stalled RTC limit our understanding of the native function, outcome, and regulation of CoV backtracking. Here, using *in vitro* RNA elongation assays, we demonstrate that the addition of nsp13 to the RTC hinders RNA elongation *in vitro* likely because its association with the RTC is unregulated. Further, we

identify a backtracking-nsp13-dependent stalling of the SARS-CoV-2 RTC prior to mismatch incorporation.

Results:

To study the nsp13s' role in SARS-CoV-2 core-RTC elongation we developed an *in vitro* primer extension assay that tracks the elongation of an RNA primer annealed to an RNA template. The RNA primer is synthesized with a 5' fluorophore so that elongation can be tracked by running samples on a denaturing urea-PAGE gel followed by fluorescent imaging. For these experiments, we designed the RNA primer-template pair with 33 nucleotides (nts) of dsRNA with 22 nts of template 5' overhang. Based on the structure of the SARS-CoV-2 backtracked RTC, 22 nts of 5' template overhang is sufficient to bind the backtracking nsp13, as it was shown to bind +8 to +14 nts of downstream template RNA (17).

Using recombinantly expressed nsp7, nsp8, and nsp12 we were able to assemble an active SARS-CoV-2 core-RTC that elongated the RNA primer to the length of the template (55 nts) (**Fig 1**). Addition of nsp13 to reactions resulted in greatly reduced extension, but RNA elongation still occurred (**Fig 1**). There was an accumulation of incomplete extension reactions producing a banding pattern of extension products. Nsp13 uses ATP hydrolysis as the energy source for helicase activity (12). To test if nsp13 was reducing RNA extension by diminishing the NTP pool in reactions we screened increasing NTP concentrations from 0.04 – 2.00 mM (**Fig 1A**). Across the NTP concentrations tested the amount of RNA extension did not increase appreciably, indicating that nsp13 was reducing primer extension via another mechanism (**Fig 1A**). To assess if nsp13 was disrupting RTC formation prior to NTP addition we altered when nsp13 was added to reactions (**Fig 1B**). No matter when nsp13 was added, even if at the same time as NTPs, it resulted in the same reduction in primer extension (**Fig 1B**).

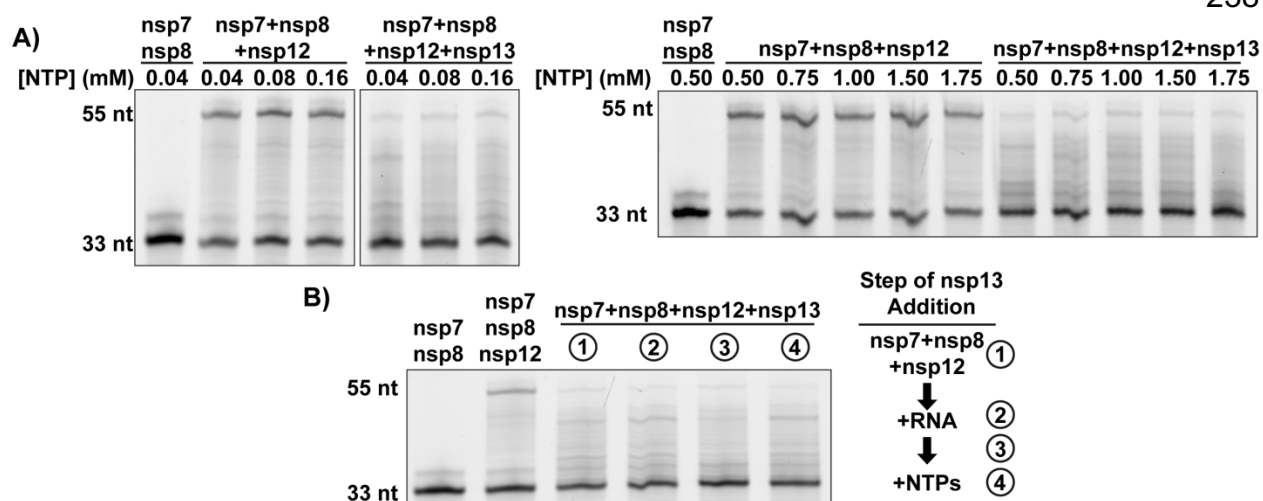


Figure 1, Nsp13 reduces *in vitro* RNA elongation: A) A 33 nt primer is extended to the length (55 nts) of the template RNA it is annealed in the presence of the SARS-CoV-2 core-RTC. Addition of nsp13 to the *in vitro* reaction reduces the amount of fully extended RNA and this reduction is independent of [NTP]. **B)** Nsp13 was added to elongation reactions at the same time as the other nsps (1), when RNA is added (2), between RNA and NTP addition (3), or with NTPs (4).

To evaluate if the backtracking nsp13 was binding 5' template RNA and hindering the RTC we tested the extension of an RNA duplex with a 5' template overhang that is too short to bind the backtracking nsp13 (**Fig 2A**). This shortened RNA duplex was extended equally in the presence or absence of nsp13, indicating that binding of nsp13 to the 5' template was reducing RTC elongation (**Fig 2A**). To evaluate if nsp13's helicase activity on the 5' template RNA was reducing RTC elongation *in vitro* we produced two ATPase knockout mutants of nsp13 (K288A, and D374A). The addition of these two ATPase knockout nsp13s to extension reactions resulted in elongation amounts equivalent, if not more, to reactions without nsp13 (**Fig 2B**). Taken together, these results indicate that nsp13 is reducing RTC elongation *in vitro* by having unregulated helicase activity on the template 5' end of RNA.

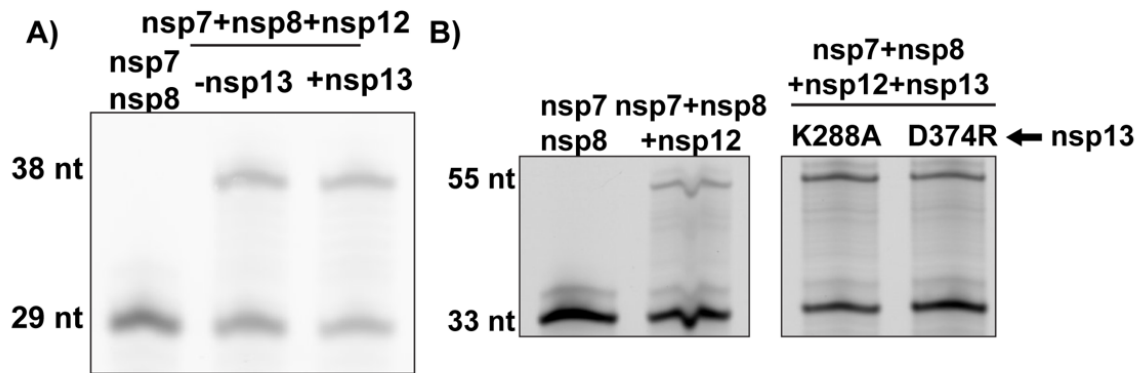


Figure 2, Nsp13's helicase activity on the 5' template RNA causes reduction in extension:
A) *In vitro* RNA elongation assay using a shorted RNA duplex (29 nt of dsRNA, 9 nt of 5' template overhang) that has a 5' overhang too short to bind the backtracking nsp13. **B)** RNA elongation assay using the long dsRNA duplex with the addition of ATPase knockout nsp13 (K288A and D374R).

We were interested in determining the role of nsp13 backtracking, or lack thereof, in CoV proofreading. We tested this by running extension reactions with only ATP, CTP and UTP added with our longer dsRNA duplex. Based off the sequence of downstream template, after the incorporation of 5 nts, the RTC would be forced to either stall or incorporate a mismatch across from a template cytosine (**Fig 3**). In the absence of nsp13, there is an accumulation of RNA products at +5, +15, and +21 nts (**Fig 3A**). These indicate that the RTC pauses prior to a misincorporation, but once misincorporation occurs it extends until the next pause. In the presence of nsp13, there is primarily an accumulation of RNA products at +5, indicating that nsp13 is promoting pausing, and possibly backtracking prior to misincorporation (**Fig 3A**). A possible explanation for why we always see a spread of RNA extension products in the reaction with nsp13 and all NTPs is that nsp13's unregulated helicase activity is causing premature abortive elongation termination, resulting in RNA's that have been elongated to various lengths.

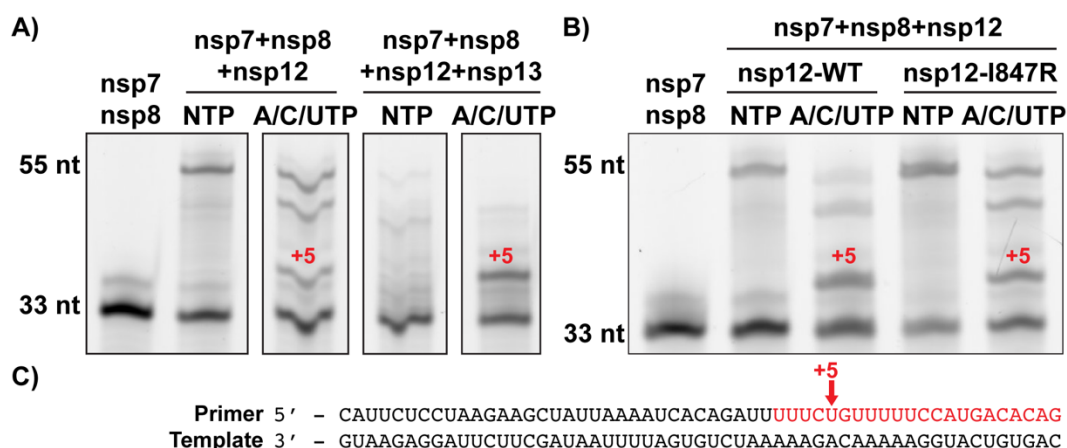


Figure 3, Nsp8_T and nsp13 promote pre-misincorporation pausing: A) *In vitro* RNA elongation assays ran without GTP present in the absence/presence of nsp13. **B)** Elongation assay comparing pre-misincorporation pausing of the core-RTC (no nsp13) between wildtype nsp12 and nsp8_T N-terminal binding disrupting mutation, nsp12-I847R. **C)** Sequence of the long RNA used in the reactions. The RNA that is synthesized during reactions is in red. The +5 pre-misincorporation pause site is indicated by the arrow and the corresponding bands are notated as “+5” in (A) and (B)

To further examine the potential role of nsp8_T in the pre-misincorporation pausing we tested SARS-CoV-2 mutant I847R that was designed to block nsp8_T's N-terminal extension binding to the RTC and dsRNA. With the mutant, we ran *in vitro* primer extension reactions lacking GTP. In these reactions we saw pre-misincorporation pausing like before (**Fig 3B**). Interestingly, reactions with the mutant nsp12 (I847R) saw an increased amount of extension past misincorporation events, indicating a reduction in pausing (**Fig 3B**). While these results are interesting, they have not been followed up on extensively, and replicate experiments are needed to validate this observation. Further, performing similar experiments on longer RNA substrates in the presence of nsp13 could better delineate nsp8_T and the backtracking nsp13's role in pausing. These results demonstrate that nsp8_T's N-terminal binding to dsRNA is important for inducing pausing of the RTC in addition to scaffolding nsp13.

Methods:

Recombinant protein expression and purification:

Recombinant SARS-CoV-2 nsp7, nsp8, and nsp12 were prepared as described previously (19).

Recombinant SARS-CoV-2 nsp13 was expressed in Rosetta 2pLysS *E. coli* cells (Novagen). 1 L cultures were grown at 37°C until they reached an OD₆₀₀ of 0.6-0.8. at which they were induced via addition of IPTG (final concentration = 500 μM) and ZnCl₂ (final concentration = 100 μM). Cultures shook at 16°C for 16 hours before cells were harvested via centrifugation and resuspended in Ni-wash buffer (25 mM HEPES, 300 mM NaCl, 30 mM imidazole, 2 mM DTT). Resuspended cells were lysed via microfluidization (Microfluidics) and lysates cleared via centrifugation followed by filtration. Protein was batch bound then purified using Ni-NTA agarose beads (Qiagen), eluting with wash buffer containing 300 mM imidazole. Protein was dialyzed overnight with the addition of 1% 3C protease (w/w) in dialysis buffer (10 mM HEPES, 300 mM NaCl, 2 mM DTT) to remove the hexa-histidine tag. The following day undigested protein was removed by flowing sample over Ni-NTA beads. Digested protein was further purified using a Superdex 200 Increase 10/300 GL column (Cytia) with SEC buffer (10 mM HEPES, 300 mM NaCl, 2 mM DTT). Fractions containing nsp13 were pooled and concentrated using ultrafiltration. Concentrated samples were aliquoted, flash frozen in liquid nitrogen and stored at -80°C. Typical yield for a 1 L expression was 2-4 mg of protein.

RNA Substrates:

RNA oligos were purchased from Integrated DNA Technologies (IDT) with 5' fluorescent tags (6-FAM) on the primer RNAs.

Long RNA duplex:

Primer: 5' CAUUCUCCUAAGAAGCUAUUAAAAUCACAGAUU 3'

Template: 3' GUAAGAGGAUUCUUCGAUAAUUUUAGUGUCUAAAAAGACAAAAGGUACUGUGAC 5'

Short RNA duplex:

Primer: 5' CAUUCUCCUAAGAAGCUAUUAAAAUCACA 3'

Template: 3' GUAAGAGGAUUCUUCGAUAAUUUUAGUGUUGGGAAAAA 5'

In vitro primer extension:

In vitro assays were performed as described previously except for the following alterations (19). To assemble complexes nsp7, nsp8, and nsp12 were combined in

reaction buffer and incubated for 15 minutes at 25°C, then RNA duplex was added, and reactions incubated for 15 minutes at 25°C, then nsp13 was added (if present) and reactions incubated for another 15 minutes, unless noted otherwise, at 25°C before the addition of NTPs. The final concentration of nsp13 in reactions was 1 μ M. Unless noted otherwise in the figure, the final concentration of NTPs was 40 μ M.

Discussion and future directions:

The CoV helicase nsp13 is a critical enzyme involved in CoV replication that has been shown to interact with the core-RTC (14,15,17). Structures of the SARS-CoV-2 RTC bound to nsp13 provided novel insight into helicase-polymerase coupling, but these structural snapshots do not provide mechanistic details into nsp13s' role and regulation in elongation and backtracking. Using an *in vitro* RNA extension assay we have established that the addition of nsp13 to the SARS-CoV-2 RTC results in reduced RNA primer extension via nsp13's helicase activity on the 5' end of template RNA. Further, we have identified that the backtracking nsp13 promotes the pausing of an active elongation complex prior to a misincorporation event.

By testing various NTP concentrations, shortened RNA duplexes, and nsp13 ATPase knockouts we identified that the helicase activity of the backtracking nsp13 (bound to nsp8_T) is likely responsible for the reduced RNA extension *in vitro*. We predict that our results are a consequence of the absence of nsp13 regulation mediated backtracking in our reductionist *in vitro* assay. In our simplified assay, we exclude countless variables that could be involved in the regulation of CoV backtracking, including other viral cofactors, enzymes, and potential host factors. Further, the assembly of a stable core-RTC composed of nsp12, nsp8_F, and the nsp7+nsp8_T heterodimer bound to RNA prior to nsp13 addition may not be representative of an elongating RTC. In support of this, recent work from our lab and others established that the N-terminal extension of nsp8_T is not required for *in vitro* RNA synthesis (19,20). We have since proposed that nsp8_T could potentially regulate other steps in CoV replication, including RTC backtracking. Further, the nsp8_T N-terminal binding mutant (nsp12-I847R) resulted in increased misincorporation and thus reduced pausing *in vitro*. Taken together, our results indicate that the *in vitro* assembly of a stalled core-RTC provides the scaffolding for nsp13 to bind the RTC and cause unregulated backtracking.

Forcing potential mismatches by excluding GTP from reactions resulted in an accumulation of RNA products that had been elongated by 5, 15, and 21 nts in the absence of nsp13. Based off the RNA primer-template sequence, these RNAs have been stalled prior to a nt misincorporation. With the addition of nsp13, we saw a strong accumulation of RNA products only at +5 nts. These results indicate that the RTC pauses prior to a misincorporation, and in the presence of nsp13, RTCs remained paused avoiding misincorporation.

Although we have determined that nsp13's helicase activity on the 5' template RNA is responsible for the reduction in RNA extension, our *in vitro* assay is unable to assess if backtracking is occurring. To test this one could use a combination of *in vitro* primer extension with 4-thio-uracil crosslinking. For this, a 4-thio-uracil could be synthesized at or near the 3' end of the primer on the annealed dsRNA substrate. This substrate could then be used for misincorporation assays where the forced misincorporation/pause is

near the primer 3' end. After primer extension, samples could be UV-crosslinked and only if backtracking occurred the 4-thio-uracil would be able to crosslink in nsp12's NTP channel. Further, nsp8 mutations could be designed to specifically disrupt binding of the backtracking nsp13. These mutants would determine if nsp13 interaction with the RTC is responsible for our reported results.

References:

1. Fehr, A.R. and Perlman, S. (2015) Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol Biol*, **1282**, 1-23.
2. Brian, D.A. and Baric, R.S. (2005) Coronavirus genome structure and replication. *Curr Top Microbiol Immunol*, **287**, 1-30.
3. van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E. and Snijder, E.J. (2008) SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. *PLoS Pathog*, **4**, e1000054.
4. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
5. Klein, S.A.-O., Cortese, M.A.-O., Winter, S.L., Wachsmuth-Melm, M.A.-O., Neufeldt, C.A.-O., Cerikan, B., Stanifer, M.A.-O., Boulant, S., Bartenschlager, R.A.-O. and Chlanda, P.A.-O. SARS-CoV-2 structure and replication characterized by in situ cryo-electron tomography.
6. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.T., Meng, X., Xia, H., Wood, J.M., Harris, L.D. *et al.* (2021) Inhibition of SARS-CoV-2 polymerase by nucleotide analogs: a single molecule perspective. *bioRxiv*.
7. Malone, B., Urakova, N., Snijder, E.J. and Campbell, E.A. (2021) Structures and functions of coronavirus replication–transcription complexes and their relevance for SARS-CoV-2 drug design. *Nature Reviews Molecular Cell Biology*.
8. Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res*, **96**, 59-126.
9. Kirchdoerfer, R.N. and Ward, A.B. (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun*, **10**, 2342.
10. Hillen, H.S., Kobic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154-156.

11. Wang, Q., Wu, J., Wang, H., Gao, Y., Liu, Q., Mu, A., Ji, W., Yan, L., Zhu, Y., Zhu, C. *et al.* (2020) Structural Basis for RNA Replication by the SARS-CoV-2 Polymerase. *Cell*, **182**, 417-428 e413.
12. Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., Zheng, L., Ming, Z., Zhang, L., Lou, Z. and Rao, Z. (2019) Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res*, **47**, 6538-6550.
13. Sommers, J.A., Loftus, L.N., Jones, M.P., 3rd, Lee, R.A., Haren, C.E., Dumm, A.J. and Brosh, R.M., Jr. (2023) Biochemical analysis of SARS-CoV-2 Nsp13 helicase implicated in COVID-19 and factors that regulate its catalytic functions. *J Biol Chem*, **299**, 102980.
14. Lehmann, K.C., Snijder, E.J., Posthuma, C.C. and Gorbalenya, A.E. (2015) What we know but do not understand about nidovirus helicases. *Virus Res*, **202**, 12-32.
15. Yan, L., Zhang, Y., Ge, J., Zheng, L., Gao, Y., Wang, T., Jia, Z., Wang, H., Huang, Y., Li, M. *et al.* (2020) Architecture of a SARS-CoV-2 mini replication and transcription complex. *Nat Commun*, **11**, 5874.
16. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatandaslar, H., Chait, B.T. *et al.* (2020) Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell*, **182**, 1560-1573 e1513.
17. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc Natl Acad Sci U S A*, **118**.
18. Abdelkareem, M., Saint-André, C., Takacs, M., Papai, G., Crucifix, C., Guo, X., Ortiz, J. and Weixlbaumer, A. (2019) Structural Basis of Transcription: RNA Polymerase Backtracking and Its Reactivation. *Mol Cell*, **75**, 298-309.e294.
19. Anderson, T.K., Hoferle, P.J., Chojnacki, K.J., Lee, K.W., Coon, J.J. and Kirchdoerfer, R.N. (2024) An alphacoronavirus polymerase structure reveals conserved replication factor functions. *Nucleic Acids Res*.
20. Campagnola, G., Govindarajan, V., Pelletier, A., Canard, B. and Peersen, O.B. (2022) The SARS-CoV nsp12 Polymerase Active Site Is Tuned for Large-Genome Replication. *J Virol*, **96**, e0067122.

Appendix 2, Development of *in vitro* RNA assays:

Authors: Thomas K. Anderson¹, Peter J. Hoferle¹, Robert N. Kirchdoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A. generated the data in Figures 1 and 2.

P.J.H. generated the data in Figure 3.

T.K.A. wrote this appendix with editing by R.N.K..

Introduction:

The coronavirus (CoV) (+) ssRNA genome encodes for 16 non-structural proteins or nsps (1). Among these nsps are several enzymes involved in the synthesis and modification of viral RNA products. Important for our work are the viral RNA-dependent RNA polymerase (nsp12), superfamily 1B helicase (nsp13), and 3'-5' exoribonuclease (nsp14). Two of these nsps (nsp12 and nsp14) require viral cofactor proteins for optimal enzymatic activity. Nsp12 forms the viral core replication-transcription complex (RTC) with cofactors nsp7 and nsp8 (2). Nsp14 binds the cofactor nsp10 to form the exonuclease complex (3).

In vitro assays are commonly used to characterize and test viral RNA enzymes. We developed *in vitro* RNA assays to evaluate nsp12's polymerase activity, nsp13's helicase activity, and nsp14's exonuclease activity. Here I will briefly describe the development of these assays for enzymes from SARS-CoV-2, and PEDV.

Methods:

Provided are generalized methods for each RNA assay. Variations in methods (i.e. reaction times, and temperatures) will be noted in the "**Results**" section for different virus complexes and assay development steps.

RNA and DNA oligos:

Oligos were ordered and synthesized from Integrated DNA Technologies (IDT). "Primer RNA" and "Helicase Probe" were manufactured with 5' fluorescent 6-FAM tags. To simplify the naming of RNA substrates through the main text, they will be referred to their following names of "Primer" and "Template" even if they aren't being used for elongation.

RNA

Primer RNA: CAUUCUCCUAAGAAGCUAUUAAAAUCACA

Template RNA w/ **DNA hairpin**:

AAAAAGGGAUGUGAUUUUAAUAGCUUCUAGGAGAAUG**GGCTAGCATTCTCTGC**
TAGCC

Template RNA: AAAAAGGGAUGUGAUUUUAAUAGCUUCUAGGAGAAUG

DNA

Helicase Probe/Trap: TTAACACGAGAGTAA

Helicase Template: TTTTTTTTTTTTTTTTTTTTTTTACTCTCGTGTTAA

In vitro primer extension assay:

Reaction buffer conditions were 10 mM Tris-Cl pH 8.0, 10 mM sodium chloride, 2 mM magnesium chloride, and 1 mM DTT. Protein final concentrations were 500 nM nsp12, 1.5 μ M nsp7 and 1.5 μ M nsp8. Duplex RNA final concentration was 250 nM. Proteins were diluted in assay buffer and then combined and incubated at 25°C for 15 min, duplex RNA was added and reactions incubated at 25°C for an additional 15 min. Reactions were initiated by the addition of NTPs to a final concentration of 40 μ M and reactions ran for 1 min (at 25°C for SARS-CoV-2, or 30°C for PEDV) before being halted by addition of two volumes of sample loading buffer (95% (v/v) formamide, 2 mM ethylenediaminetetraacetic acid (EDTA) and 0.75 mM bromophenol blue). Samples were heated at 95°C then analyzed using denaturing urea-PAGE (8 M urea, 15% polyacrylamide) run in 1 \times TBE (89 mM Tris-Cl pH 8.3, 89 mM boric acid, 2 mM EDTA).

Gels were imaged using a Typhoon FLA 9200 (GE Healthcare) to identify fluorescein signals. Extension could be quantified using ImageJ.

In vitro exonuclease assay:

Proteins and RNA (ss or ds) were combined at final concentrations of nsp10 - 500 nM, nsp14 - 125 nM, and RNA - 250 nM. Buffer conditions were 10 mM Tris, 10 mM NaCl, 2 mM MgCl₂, and 2 mM DTT. Proteins were diluted in reaction buffer after which proteins and RNA were combined and incubated at 30°C for 30 minutes. Reactions were halted by the addition of two reaction volumes of denaturing loading buffer (95% (v/v) formamide, 2 mM EDTA, and 0.75 mM bromophenol blue). Samples were then analyzed following an identical protocol as the “*In vitro* primer extension assay”.

In vitro helicase assay:

Reaction conditions were 10 mM Tris, 10 mM NaCl, 2 mM MgCl₂, and 2 mM DTT. Reactions were initiated by combining nsp13 (500 nM), duplex DNA (250 nM), trap DNA (1.25 μM), and ATP (500 μM). After combining reactions were incubated for 15 minutes at 30°C after which reactions were put on ice. Reaction samples were mixed with 10X helicase assay loading buffer (10 mM TRIS, 1 mM EDTA, 50% glycerol (v/v), 2% SDS, and 0.75 mM bromophenol blue) then ran on a 6% native-PAGE gel in 1X TBE. Prior to running the gel, gel rig, and running buffer were pre-chilled on ice, and gels were run in an ice bath. Gels were imaged using a Typhoon FLA 9200 (GE healthcare) to identify fluorescein signals.

Results:

Polymerase activity assay:

To assemble an active SARS-CoV-2 polymerase complex we combined recombinantly expressed nsp7, nsp8, and nsp12 with a short RNA duplex and added NTPs. For the duplex, the primer (Primer RNA) had a 5' fluorescent tag to easily track elongation, and the template had a 3' short DNA hairpin (Template RNA w/ DNA hairpin) to promote correct binding of the duplex to the polymerase. Based on prior publications we assembled the RTC+RNA in a stepwise manner, first allowing the nsps to form the core-RTC prior to the addition of RNA. We tested extension of RNA over a time course of 1 hour and observed ~50% extension by the first minute (**Fig 1A**). At later time points (>3 minutes) extension extended 1 nucleotide (nt) past the template length (**Fig 1A**). Moving forward we ran SARS-CoV-2 primer extensions for 1 minute to avoid excess polymerase activity. Further, there was a population of un-extended RNA that persisted up to 1 hour (**Fig 1A**). This indicated that RNA was not being extended. Reasons for this could include improper RNA binding, lack of RNA binding, or inactive complexes. To simplify our RNA oligo design, we compared extension using RNA templates with or without the 3' DNA hairpin (**Fig 1B**). We did not see an appreciable improvement between the two substrates, so we did not include the hairpin in future work.

For the PEDV core-RTC we initially tested using a matching protocol to the SARS-CoV-2 assay over a 10-minute time course (**Fig 1C**). We observed less polymerase activity for PEDV than SARS-CoV-2. Extension efficiency improved when running reactions at a slightly higher temperature of 30°C compared to SARS-CoV-2 reactions being run at 25°C (**Fig 1D**).

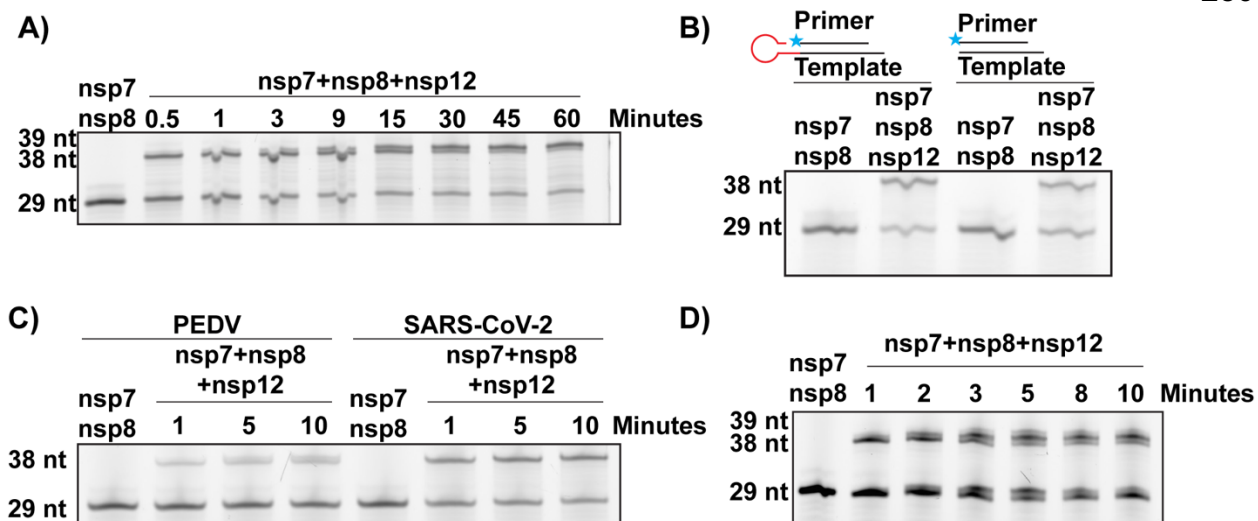


Figure 1, *in vitro* primer extension: **A)** Primer extension time course for 0.5-60 minutes for the SARS-CoV-2 core-RTC. **B)** Comparison of extension by the SARS-CoV-2 core-RTC on duplex RNAs with or without a 3' DNA hairpin (red) on the template. **C)** Primer extension time course from 1-10 minutes for the PEDV core-RTC. SARS-CoV-2 was included as a positive control. These reactions were incubated at 25°C. **D)** Primer extension for PEDV from 1-10 minutes with reactions that were incubated at 30°C.

Exonuclease activity assay:

To test the activity of recombinantly expressed exonuclease complexes (nsp10 + nsp14) we would combine our proteins with fluorescently tagged ss (Primer RNA) or dsRNA substrates (Primer RNA + Template RNA) and tracked degradation by denaturing urea-PAGE. We designed the substrates to have the fluorescent tag on the 5' end so that nsp14's 3'-5' exonuclease activity could be tracked along the whole substrate. We assessed exonuclease activity with or without the metal ion cofactor Mg^{2+} and observed Mg^{2+} dependent exonuclease activity (**Fig 2A**). We consistently had reduced activity on dsRNA compared to ssRNA. In an attempt to improve activity, we added DNA on the template 3' end to prevent degradation of template (Template RNA w/ DNA hairpin) and promote activity on the tagged primer (**Fig 2B**). Although the activity on dsRNA improved with the template 3' DNA present, the SARS-CoV-2 exonuclease still displayed a preference for ssRNA. To determine the amount of nsp10 required for optimal exonuclease activity we titrated nsp10 into exonuclease reactions and observed that a nsp10 : nsp14 ratio of 4 : 1 resulted in optimal activity (**Fig 2C**). Using our established assay, we tested the activity of the PEDV exonuclease complex at 30°C and 37°C and observed exonuclease activity, interestingly with a preference for dsRNA over ssRNA (**Fig 2D**).

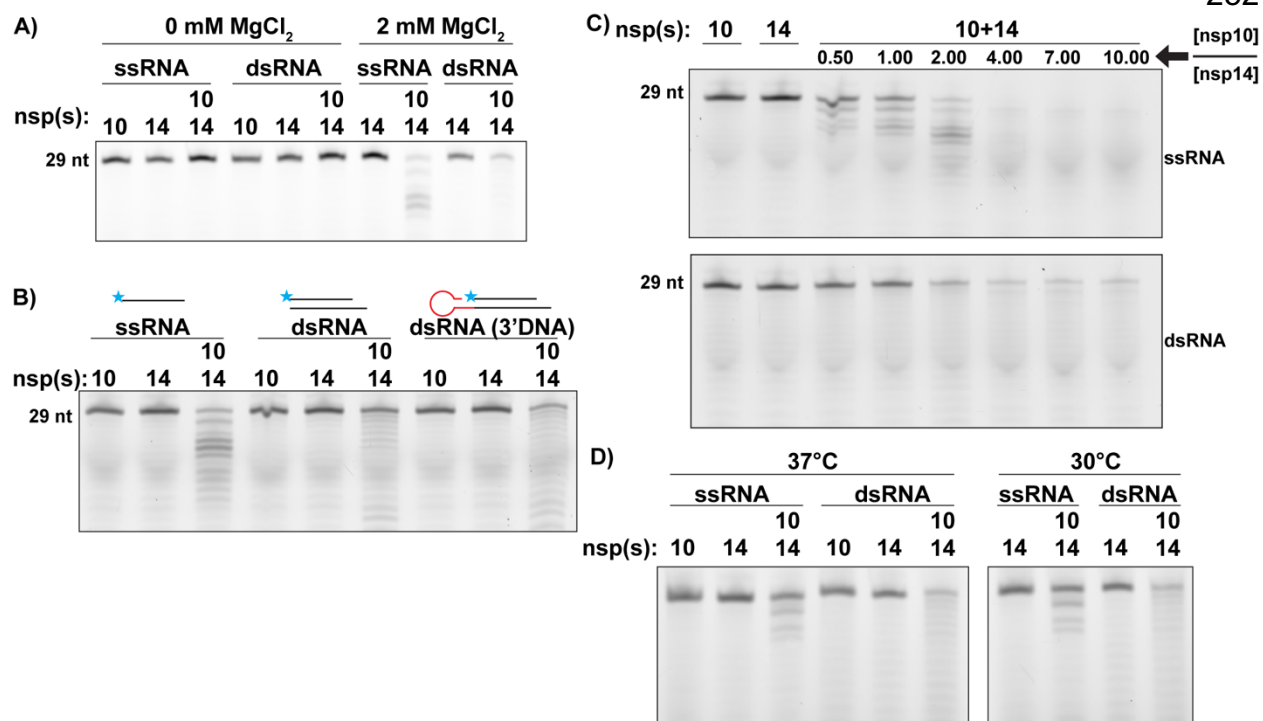


Figure 2, exonuclease assay: **A)** Exonuclease assay comparing activity of SARS-CoV-2 nsp10 and nsp14 on ss or dsRNA in the absence or presence of MgCl₂. **B)** Comparison of SARS-CoV-2 exonuclease complex activity on dsRNA duplexes with or DNA (red) on the template 3' end to promote primer degradation. **C)** Titration of nsp10 into SARS-CoV-2 exonuclease degradation reactions to find optimal ratio for nsp14 activation. **D)** Activity of the PEDV exonuclease complex at 37°C and 30 °C on ss and dsRNA.

Helicase activity assay:

To assess the helicase activity of nsp13 we designed a DNA unwinding assay with a fluorescently labelled DNA probe (Helicase Probe) that could be examined via native-PAGE. To prevent the reannealing of template DNA to the probe after unwinding, an unlabeled DNA trap (same sequence as the probe) was added to reactions. To screen how much excess trap is needed to prevent reannealing we heat-shocked samples of duplex + trap at increasing amounts of trap and found that 5x excess of trap is sufficient (**Fig 3A**). Initial helicase assays found DNA unwinding in the presence of duplex DNA, ATP, and trap (no nsp13 added) (**Fig 3B**). We predicted that high concentrations of ATP were promoting unwinding of the low- T_m DNA duplex. To reduce unwanted unwinding, we tested various concentrations of ATP +/- nsp13 and found 0.5 mM ATP to both promote nsp13's helicase activity while reducing unwinding in nsp13s absence (**Fig 3B**). Under matching conditions this assay works to assess PEDV nsp13 helicase activity (**Fig 3C**).

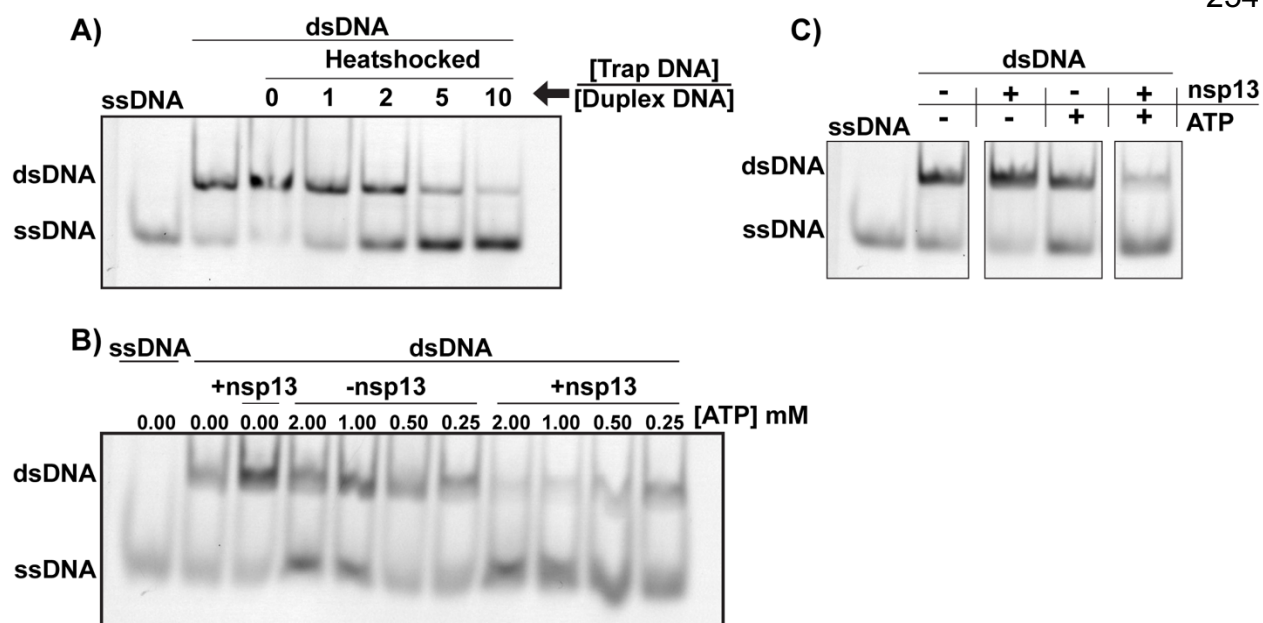


Figure 3, helicase assays: **A)** Native-PAGE gel assessing re-annealing of DNA probe and template after heat shock at 95°C for 5 minutes with increasing amounts of trap DNA oligomer present. **B)** Comparison of DNA unwinding at decreasing concentrations of ATP in the absence or presence of SARS-CoV-2 nsp13. **C)** Helicase assay assessing the activity of PEDV nsp13.

References:

1. Snijder, E.J., Decroly, E. and Ziebuhr, J. (2016) The Nonstructural Proteins Directing Coronavirus RNA Synthesis and Processing. *Adv Virus Res*, **96**, 59-126.
2. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and 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.
3. Ma, Z., Pourfarjam, Y. and Kim, I.K. (2021) Reconstitution and functional characterization of SARS-CoV-2 proofreading complex.

Appendix 3: In depth protocols used in the completion of thesis work

Authors: Thomas K. Anderson¹, Peter J. Hoferle¹, Robert N. Kirchoerfer¹

Affiliations:

¹Department of Biochemistry, Institute for Molecular Virology, Center for Quantitative Cell Imaging, University of Wisconsin-Madison, Madison, WI 53706, USA.

Author contributions:

T.K.A., P.J.H. and R.N.K. developed the following protocols through collaboration and discussion.

T.K.A. assembled this appendix with editing by R.N.K..

Introduction:

Through the completion of this dissertation several protocols were developed and used for the cloning, expression, purification, and analysis of recombinant CoV proteins. Here, I will provide detailed sample protocols, providing important information for their execution.

T4 ligation independent cloning:**PCR amplification of gene of interest:**

For each gene set up rxn:

16.4 μ L	Water	
5 μ L	5X Buffer	1X
1 μ L	DMSO	4%
0.5 μ L	10mM Each dNTPs	200 μ M each
1 μ L	10 μ M Fwd primer	0.4 μ M
1 μ L	10 μ M Rvs primer	0.4 μ M
0.4 μ L	Template DNA (Genescript) (~100 ng)	
0.4 μ L	Phusion Polymerase	0.032 U/ μ L

Thermocycle Parameters:

98°C for 30 sec	Denature
98°C for 15 sec	*18 Cycling amplification
60°C for 20 sec	(Tm – 5C)
72°C for 30 sec (15-30s/kb)	
72°C for 60sec (30-60s/kb)	Final extension
10°C hold	

- Note: Extension times are Phusion polymerase specific. If using different DNA polymerase, change parameters.

Gel Extract Genes:

1. Prepare 1% agarose gel (30 mL fresh 1X TAE, 0.3 g of agarose, melt in microwave ~1min, add 1.5 μ L of Gel-Red)
2. Add 5 uL of 6x loading buffer to each sample
3. Load 15 uL of sample to each well, and 2 uL of 1kb ladder to 1 well
4. Run gel (I often run gels of this type for 25 minutes at 130V, this is sufficient to get a good band spread and doesn't take long to run).
5. Gel purification: I have found that the Machery-Nagel gel extraction works the best (much better than Zymo).
6. Elute with 20 uL of diluted elution buffer (1:10 dilution from what is supplied by kit)

BfuAI digestion of pET46-ccdB plasmid:

X uL	MilliQ water	
2.5 uL	10x NEBuffer 3.1	1x
X uL	Plasmid DNA (pET46ccdB)	0.5 – 1ug
1 uL	BfuAI	2U
25 uL	Final Vol	

1. Pipette up and down, incubate at 37C for 14 hr
2. Gel extract product with Machery Nagel kit, elute with 50 μ L of diluted elution buffer

T4 DNA Pol Treat:

47 uL	BfuAI-treated vector DNA
6 uL	NEB buffer 2.1
6 uL	25 mM dTTP
1.2 uL	T4 DNA Pol

Insert:

15.6 uL purified PCR insert
2 uL NEB buffer 2.1
2 uL 25 mM dATP
0.4 uL T4 DNA Pol

1. Incubate at RT for 30 min, heat inactivate T4 DNA polymerase at 75°C for 20 min, dilute Vector DNA to 2.5 ng/uL with elution buffer

- Notes:

- BfuAI is specific for this plasmid.
- It is critical for the LIC that the insert and plasmid are designed so that they can be treated with Watson-Crick base dNTPs (i.e. insert is treated with dATP, plasmid is treated with dTTP) making a perfect overhang for one another.

LIC:

1. Mix 1 uL of diluted vector, with 2 uL of T4 treated insert. Incubate at RT for 4 min
2. Add 1 uL of 25 mM EDTA, incubate at RT for 5 min
3. Use 2 uL of the mixture to transform 50 uL of TOP10 E. coli cells.
4. Transformation: 2 minutes on ice, 45 seconds at 42C, 2 minutes on ice then recover.
5. Recover in 200 uL of LB shaking for 1 hr.
6. Plate on LB+agar plates with plasmid specific antibiotic resistance.

Lightning mutagenesis:

The purpose of this protocol is to mutagenize plasmid DNAs. It is based on the methods described for the Agilent QuickChange Lightning Multi Kit.

Primer design: This type of mutagenesis only requires the use of 1 primer. The primer should be 25-45 nts long, with an equal amount of sequence on either side of the mutation. The T_m of the primer should be $>75^\circ\text{C}$ and end with 1 or more G/Cs at the 3'.

1. Phosphorylate primers (20 μL)

13.5 μL Water
 2 μL 10X PNK Buffer 1X
 2 μL 10 mM ATP 1 mM
 2 μL 100 μM Primer1 10 μM
 0.5 μL PNK (NEB) 5 U

Incubate at 37 $^\circ\text{C}$ for 30 minutes

2. Amplify and Ligate (25 μL)

10.1 μL Water
 5 μL 5X Q5 Reaction Buffer 1X
 5 μL GC Enhancer 1X
 0.5 μL 10mM dNTPs 200 μM
 0.5 μL 50mM NAD+ 1 mM
 2.4 μL 10 μM each primer 0.5 μM each
 0.5 μL Template DNA <1,000 ng
 0.5 μL HiFi Taq DNA Ligase
 0.5 μL Q5 HiFi DNA Polymerase

Q5 Thermocycler parameters

95 $^\circ\text{C}$ 0:30
 95 $^\circ\text{C}$ 0:10 |*30 cycles
 60 $^\circ\text{C}$ 0:20 |
 65 $^\circ\text{C}$ 5:00 |
 65 $^\circ\text{C}$ 10:00
 10 $^\circ\text{C}$ hold

After PCR: Remove template DNA

Add 1 μL DpnI to each 25 μL PCR reaction

Incubate at 37 $^\circ\text{C}$ for 1 hour

Transform the DNA into E. coli

1. Thaw chemically competent TOP10 E. coli on ice (10-15 min)
2. Add 4 μL of DpnI-digested PCR product to 50 μL of E. coli
3. Incubate on ice for 2 minutes
4. Heat shock at 42 $^\circ\text{C}$ for 45 sec
5. Return to ice for 2 min
6. Add 150 μL LB and incubate at 37 $^\circ\text{C}$ with shaking for 1 hour
7. Pre-warm a LB+Amp agar plate
 - o Plasmid specific antibiotic
8. Plate 200 μL of transformed culture onto the agar plate and incubate at 37 $^\circ\text{C}$ overnight

Recombinant expression and purification of coronavirus cofactor proteins:

This protocol can be used and adapted for the purification of SARS-CoV-2 or PEDV nsp7, nsp8, or nsp10. Each of these proteins are very stable expressers, often making 10s of mgs per a 1-L culture. Note: PEDV nsp10 is not as good of an expresser but this protocol still works.

Ni Bind/Wash Buffer

HEPES 25 mM or TRIS 10 mM

NaCl 300 mM

Imidazole 30 mM

pH 7.5 for HEPES, 8 for TRIS

DTT 2 mM

Ni Elution Buffer

HEPES 25 mM or TRIS 10 mM

NaCl 300 mM

Imidazole 300 mM

pH 7.5 for HEPES, 8 for TRIS

DTT 2 mM

Dialysis Buffer

HEPES 25 mM or TRIS 10 mM

NaCl 300 mM

pH 7.5 for HEPES, 8 for TRIS

DTT 2 mM

SEC Buffer

HEPES 25 mM or TRIS 10 mM

NaCl 300 mM

pH 7.5 for HEPES, 8 for TRIS

DTT 2 mM

Expression:

1. Transform 0.5 μ L of plasmid (~100 ng) into 50 μ L of Rosetta2 pLysS cells (recover with 150 μ L of LB)
 - a. Transformation: 2 min on ice, 45 sec at 42C, 2 min on ice, recover
2. Directly inoculate an overnight 100 mL LB+antibiotic starter culture from recovered cells
 - a. Inoculated culture can sit on benchtop until the end of the day, but you want to wait to put it in the shaker until the afternoon, so it doesn't go overnight for >16 hrs.
3. The next morning: Use 10 mL of the starter culture to inoculate 1 L of LB+antibiotic and incubate at 37°C, 220 rpm
 - a. NOTE: use baffle flasks for 1L cultures to promote aerobic cell growth
4. Allow culture to grow to OD₆₀₀ = 0.6-0.8
 - Once there, induce 1 L of culture with 1 mL of 0.5 M IPTG
5. Incubate the cultures at 16°C overnight, 220 rpm
6. The next morning: spin down at 3,500 x g for 15 min, discard the supernatant
 - a. The pellet can be prepared immediately or frozen and stored at -80°C.

7. Resuspend the pellets in 40 mL of Ni Bind/Wash Buffer per 1 L culture.
 - a. If starting with a frozen pellet, let the pellet warm up for a bit (30 min) before trying to resuspend
8. Lyse cells in the microfluidizer
 - a. 20,000 psi. The entirety of the sample should pass through the shearing chamber at least twice.
9. Clear each lysate by centrifugation at 25,000 x g for 30 min
10. Filter the supernatant (0.45 µm) and bind in batch to 2 mL of Ni-NTA beads for at least 30 min (up to 2 hours)
 - a. Batch bind: gentle rotation at 4°C is preferred. Some proteins dislike the cold (i.e. CoV nsp15) and should batch bind at 25°C.
 - b. If you have issues with low yields and E. coli contaminating proteins, you can use 1 mL of beads.
11. Collect beads by centrifugation (500 x g for 5 min) and remove the supernatant before loading onto a gravity column
12. Wash the beads with 2*20 mL Bind/Wash buffer
13. Elute with 5*1 CV of Elution buffer
14. Run fractions on SDS-PAGE from the various steps (supernatant, pellet, washes, elutions)
15. Pool Ni elutions into a single tube. Measure the A₂₈₀ of the pooled fractions. Calculate mg/mL concentration
16. Add 1% (w/w) of TEV protease to all of sample
17. Transfer each protein to dialysis tubing (3.5 kDa MW cutoff) and dialyze into Dialysis Buffer overnight at 4°C
18. For the cleaved portions
 - a. Flow the dialyzed protein over 2 mL of equilibrated Ni-NTA collecting the flow through (cleaved protein).
 - b. Wash the column with 6 mL of Dialysis Buffer adding this protein to the flow through
 - c. Elute from the column into a separate tube (digested but uncleaved protein)
19. For the cleaved protein (Ni flow through) and the undigested portion of the proteins, concentrate the protein to < 500 µL
20. Hard spin each sample (5min at 16,000g)
21. Load each sample onto a Superdex200 10/300 Increase column equilibrated in SEC Buffer
 - a. Columns can be overloaded, I recommend not running more than 10 mg at a time, at least on this column.
22. Pool fractions containing your protein of interest and concentrate using ultrafiltration. These cofactors can be concentrated to >10 mg/mL. Once concentrated aliquot (20-30 µL aliquots are good for our in vitro assays) and flash freeze in liquid nitrogen. Store at -80°C.

Preparation of P0 and P1 baculovirus stocks:Transfection for P0 generation:

1. Seed each well of a 6 well plate with 1.5×10^6 Sf9 insect cells in 1.5 mL ESF 921 media. Allow 20 min undisturbed to attach.
2. Mix the Cellfectin II transfection reagent well before use via inversion
3. Transfection Mixture (combine in order):
 - 1 mL ESF 921 media
 - Bacmid DNA: 1 μ g
 - Cellfectin II: 10 μ L
 - Make sure to include -DNA control.
4. Mix well by inversion, incubate at 25°C for 20 min
5. Discard original medium from plate and add transfection mixture dropwise (and gently!) to the cells surface, ensure to add drops to the full surface area. Rock back and forth gently. Incubate at 27°C for 4 hours. Then add 3 mL of fresh media, and tape 3 sides of the plate. Incubate for 5 days at 27°C.
6. Harvest virus (supernatant) after 5 days via centrifugation. 1000 x g for 5 min
 - a. Move supernatant to a clean tube after centrifugation, store at 4°C. Can be stored for a few weeks.

For a 12 well plate, modify protocol as follows: seed 0.5×10^6 cells/well in 1 mL media.

1. Transfection mixture:
 - a. 200 μ L media
 - b. Bacmid: 300 ng
 - c. 3 μ L Cellfectin II
2. Then when adding fresh media add 1 mL

Generation of P1 Stock:

1. Seed a 12 mL adherent culture with 6×10^6 Sf9 cells and infect with 0.5 mL of P0 virus
2. Incubate for 5 days at 27°C
3. Harvest P1 virus by centrifugation (1,000 x g 5 min). Can be stored at 4°C for 6-12 months.

Recombinant expression and purification of SARS-CoV-2 and PEDV nsp12:

This protocol is to prepare SARS-CoV-2 or PEDV nsp12 from baculovirus infected cells using strep chromatography.

Strep Wash Buffer

HEPES 25 mM
NaCl 300 mM
MgCl₂ 1 mM

pH→7.4 with HCl

0.45 µm Filter

DTT 2 mM

Strep Elution Buffer (10 mL)

10 mL Strep Wash 1X
5.2 mg Desthiobiotin 2.5mM

409 SEC Buffer (250 mL)

HEPES 25 mM
NaCl 300 mM
50 µL 0.5 M MgCl₂ 100 µM
0.5 mL 1 M TCEP 2 mM

pH→7.4 with NaOH

(prepare SEC buffer the day you use it)

1. Day 0: Start amplifying Sf21 cells (250 mL at 1e6 cells/mL) and start P2 virus stock (50 mL of Sf9 cells at 1e6 cells/mL + 0.4 mL of P1 virus stock (protocol for P1 generation above))
2. Day 3: Expand Sf21 cells to 1 L at 0.8-1.0e6 cells/mL
3. Day 4: Infect Sf21 cells with 10 mL of P2 virus stock
 - a. Spin down the P2 stock prior to use (1,000 x g for 10 min at 4°C)
 - b. Infect 1*1 L cultures of Sf21 cells with 10 mL of P2 baculovirus stock (supposed moi of 5). Incubate at 27°C.
4. Day 6: Spin down 1 L culture (1,000 x g for 10 minutes, 4°C)
5. Discard the supernatant
6. Resuspend cells in 100 mL Wash buffer and add 143 µL BioLock/L of cell culture
7. Lyse the cells using the microfluidizer at 20,000 psi
8. Clear lysates by centrifugation at 25,000 x g at 4°C for 30 minutes
9. Filter lysates (0.45 µm filter)
10. Add 2 mL of Streptactin Superflow Agarose to cleared lysates and incubate with agitation for 30 min at 4°C
11. Collect resin by gentle centrifugation (500 x g for 5 min) and load gravity column
12. Wash twice with 20 mL of Wash buffer
13. Elute slowly with 10 mL of Elution Buffer,
14. Run samples on SDS-PAGE
15. Clean beads by: water wash, 3 CV of 0.5 M NaOH, 3 water washes, then wash out of column and store at 4°C
 - a. NOTE: strep beads don't last very long, so you can usually reuse them a couple times for like 2 weeks but don't store for longer than that.
16. If SDS-PAGE looks good, can-do SEC same day or store sample at 4°C O/N.

Recombinant expression and purification of IBV nsp12:Ni Bind/Wash Buffer (500 mL)

HEPES	25 mM
NaCl	300 mM
Imidazole	30 mM
MgCl ₂	1 mM
pH 7.5	
DTT	2 mM

Ni Elution Buffer (100 mL)

HEPES	10 mM
NaCl	300 mM
Imidazole	300 mM
MgCl ₂	1 mM
pH 7.5	
DTT	2 mM

Dialysis Buffer (1L)

HEPES	10 mM
NaCl	300 mM
MgCl ₂	2 mM
pH 7.5	
DTT	2mM

SEC Buffer (500 mL)

HEPES	25 mM
NaCl	300 mM
MgCl ₂	100 μM
pH 7.5	
1 mL 0.5 M TCEP	2 mM

Expression:

1. Transform 0.5 μL of plasmid (~100 ng) into 50 uL Rosetta2 pLysS cells (recover with 150 uL of LB)
2. Directly inoculate an overnight 100 mL LB+antibiotics starter culture from recovered cells
3. Use 10 mL of the starter culture to inoculate 1 L of LB+antibiotics and incubate at 37°C, 220 rpm
4. Allow culture to grow to OD₆₀₀ = 0.6-0.8
 - Induce 1 L of culture with 1 mL of 0.5 M IPTG, Incubate the cultures at 16°C overnight
5. Spin down at 3,500 x g for 15 min, discard the supernatant.
6. Resuspend the pellets in 40 mL of Ni Bind/Wash Buffer per 1 L culture. Lyse cells in the microfluidizer
7. Clear each lysate by centrifugation at 25,000 x g for 30 min
8. Filter the supernatant (0.45 μm) and bind in batch to 2 mL of Ni-NTA beads for at least 30 min (up to 2 hours)
9. Collect beads by centrifugation (500 x g for 5 min) before loading onto a gravity column (skinny column)
10. Wash the beads with 2*20 mL Bind/Wash buffers
23. Elute with 5*1 CV of Elution buffer
24. Run fractions on SDS-PAGE
25. Pool Ni elutions into a single tube. Measure the A₂₈₀ of the pooled fractions. Calculate mg/mL concentration
26. Add 1% (w/w) of TEV protease to all of each sample
27. Transfer each protein to dialysis tubing and dialyze into Dialysis Buffer overnight at 4°C
28. For the cleaved portions
 - a. Flow the dialyzed protein over 2 mL of equilibrated Ni-NTA collecting the flow through (cleaved protein).

- b. Wash the column with 6 mL of Dialysis Buffer adding this protein to the flow through
 - c. Elute from the column into a separate tube (digested but uncleaved protein)
29. For the cleaved protein (Ni flow through) and the undigested portion of the proteins, concentrate the protein to < 500 μ L
 30. Hard spin each sample (5min at 16,000 x g)
 31. Load each sample onto a Superdex200 Increase column equilibrated in SEC Buffer

Recombinant expression and purification of coronavirus nsp14:Ni Bind/Wash Buffer (500 mL)

HEPES	25 mM
NaCl	300 mM
Imidazole	30 mM
0.5 M MgCl ₂	2 mM
pH→7.5	
DTT	2mM

Ni Elution Buffer (100 mL)

HEPES	25 mM
NaCl	300 mM
Imidazole	300 mM
pH→7.5	
DTT	2 mM

Dialysis Buffer (1 L)

HEPES	25 mM
NaCl	300 mM
pH → 7.5	
DTT	2 mM

SEC Buffer (500 mL)

HEPES	25 mM
NaCl	300 mM
pH→7.5	
DTT	2 mM

Expression:

1. Transform 0.5 μ L of each plasmid (~100 ng) into 50 μ L C41(DE3) pLysS cells (recover with 150 μ L of LB)
 - a. Transformation: 2 min on ice, 45 sec at 42°C, 2 min on ice, recover
2. Directly inoculate an overnight 100 mL LB+antibiotic starter culture from recovered cells
 - a. Inoculated culture can sit on benchtop until the end of the day, but you want to wait to put it in the shaker until the afternoon, so it doesn't go overnight for >16 hrs.
3. The next morning: Use 10 mL of the starter culture to inoculate 1 L of LB+antibiotic and incubate at 37°C, 220 rpm
 - a. NOTE: use baffle flasks for 1 L cultures to promote aerobic cell growth
4. Allow culture to grow to OD₆₀₀ = 0.6-0.8
 - Once there, induce 1 L of culture with 1 mL of 0.5 M IPTG
5. Incubate the cultures at 16°C overnight, 220 rpm
6. The next morning: Spin down at 3,500 x g for 15 min, discard the supernatant
 - a. The pellet can now either be prepared immediately or frozen and stored at -80°C for a later date.
7. Resuspend the pellets in 40 mL of Ni Bind/Wash Buffer per 1 L culture.
 - a. If starting with a frozen pellet, let the pellet warm up for a bit (30 min) before trying to resuspend
8. Lyse cells in the microfluidizer
 - a. 20,000 psi. The entirety of the sample should pass through the shearing chamber at least twice. For 40 mL of sample with the Kirchdoerfer lab microfluidizer, this requires pouring sample back through 3 times.
9. Clear each lysate by centrifugation at 25,000 x g for 30 min
10. Filter the supernatant (0.45 μ m) and bind in batch to 2 mL of Ni-NTA beads for at least 30 min (up to 2 hours)
 - a. If you have issues with low yields and e. coli contaminating proteins you can use 1 mL of beads.

11. Collect beads by centrifugation (500x g for 5 min) and remove the supernatant before loading onto a gravity column
12. Wash the beads with 2*20 mL Bind/Wash buffers
13. Elute with 5*1 CV of Elution buffer
 14. Run fractions on SDS-PAGE from the various steps (supernatant, pellet, washes, elution)
15. Pool Ni elutions into a single tube. Measure the A_{280} of the pooled fractions. Calculate mg/mL concentration
16. Add 1% (w/w) of TEV protease to all of sample
17. Transfer each protein to dialysis tubing (3.5 kDa MW cutoff) and dialyze into Dialysis Buffer overnight at 4°C
18. For the cleaved portions
 - a. Flow the dialyzed protein over 2 mL of equilibrated Ni-NTA collecting the flow through (cleaved protein).
 - b. Wash the column with 6 mL of Dialysis Buffer adding this protein to the flow through
 - c. Elute from the column into a separate tube (digested but uncleaved protein)
19. For the cleaved protein (Ni flow through) and the undigested portion of the proteins, concentrate the protein to < 500 μ L
20. Hard spin each sample (5min at 16,000 x g)
21. Load each sample onto a Superdex200 Increase 10/300 column equilibrated in SEC Buffer
 - a. Columns can be overloaded, I recommend not running more than 10mg at a time, at least on this column.
22. Pool fractions containing your protein of interest and concentrate using ultrafiltration.

Recombinant expression and purification of coronavirus nsp13:Ni Bind/Wash Buffer (500 mL)

HEPES	25 mM
NaCl	300 mM
1Imidazole	30 mM
pH→7.4	
DTT	2 mM

Ni Elution Buffer (100 mL)

HEPES	25 mM
NaCl	300 mM
Imidazole	300 mM
pH→7.4	
DTT	2 mM

Dialysis Buffer (1 L)

HEPES	10 mM
NaCl	300 mM
pH→7.4	
DTT	2 mM

SEC Buffer (500 mL)

HEPES	10 mM
NaCl	300 mM
pH→7.4	
DTT	2 mM

Expression:

- 1 Transform 0.5 μ L of each plasmid (~100 ng) into 50 μ L Rosetta 2pLys cells (recover with 150 μ L of LB)
- 2 Directly inoculate an overnight 100mL LB+antibiotic starter culture from recovered cells
- 3 Use 10 mL of the starter culture to inoculate 1 L of LB+antibiotic and incubate at 37°C, 220 rpm
- 4 Allow culture to grow to OD₆₀₀ = 0.8
 - a. Induce 1 L of culture with 1 mL of 0.5M IPTG and an additional 200 μ L of 0.5 M ZnCl₂
- 5 Incubate the cultures at 16°C overnight
- 6 Spin down at 3,500 x g for 15 min, discard the supernatant.
- 7 Resuspend the pellets in 40 mL of Ni Bind/Wash Buffer per 1000 mL culture. Lyse cells in the microfluidizer
- 8 Clear each lysate by centrifugation at 25,000 x g for 30 min
- 9 Filter the supernatant (0.45 μ M) and bind in batch to 2mL of Ni-NTA beads for at least 30 min (up to 2 hours)
 - a. Preferred batch binding is gentle rotation at 4°C
- 10 Collect beads by centrifugation (500 x g for 5 min) before loading onto a gravity column (skinny column)
- 11 Wash the beads with 2*20 mL Bind/Wash buffers
- 12 Elute with 5*1 CV of Elution buffer
- 13 Run fractions on SDS-PAGE
- 14 Pool Ni elutions into a single tube. Measure the A₂₈₀ of the pooled fractions. Calculate mg/mL concentration
- 15 Add 1% 3C protease (w/w) and incubate in dialysis at 4°C O/N
- 16 Inverse bind protein over 2 mL Ni-NTA
- 17 concentrate the protein to < 500 μ L
- 18 Hard spin each sample (5min at 16,000 x g)
- 19 Load each sample onto a Superdex200 Increase column equilibrated in SEC buffer

In vitro* primer extension assay example protocol:*10x HEPES Buffer:**

HEPES 250 mM
 NaCl 500 mM
 MgCl₂*6H₂O 200 mM
 pH to 7.5, syringe filter

RNA Loading Buffer:

Formamide 9.5 mL
 Bromophenol blue
 EDTA (Final: 2 mM)

RNA Binding Buffer:

KCl 100 mM
 HEPES 100 mM
 MgCl₂*6H₂O 20 mM
 pH 7.4 and filter

NOTE: Several buffers can work for primer extension depending on the CoV polymerase being studied. This includes using pH 7.5 (HEPES) or pH 8 (TRIS) buffering agents with added NaCl, KCl, Na-Glutamate, or K-Glutamate (100-1000 mM in 10X stock) and MgCl₂ (20 mM in 10X stock).

Reagent	[Start]	Final Conc.
DEPC H ₂ O	water	
10x Buffer	10x	1x
DTT	10 mM	1 mM
Nsp12	10 μM	0.5 μM
Nsp 7	30 μM	1.5 μM
Nsp8	30 μM	1.5 μM
Primer:Temp	5 μm	0.25 μM
NTPs	800 μM	40 μM

Reaction Conditions:

Typical reaction volume = 20 μL

Annealing RNA primer and template:

Want a 1:1.2 ratio of primer;template

Combine RNA with RNA binding buffer (1 μL per 20 μL total volume) and DEPC treated water.

Heat at 95°C for 5 minutes, turn heat block off and allow to cool in cooling heat block for 1 hr. Move to RT for 15 minutes. Samples can now be stored at -20°C.

Reaction Protocol:

Dilute protein in 1X reaction buffer

Mix up Water, buffer, DTT and nsp7+8+12. Gently mix and then incubate at 25°C for 15 minutes

Add RNA, gently mix, and incubate at 25°C for 15 minutes

Add NTPs and incubate at 25°C for 1 minute

To quench reaction, mix 5 μL of sample with 10 μL of RNA Loading Buffer.

Gel:

Combine 4.8 g of urea with 0.5 mL of 10X TBE, 5 mL of 30% Acrylimide:Bis-Acrylimide 19:1, and DEPC treated water (to 10 mL), dissolve the urea. Clean mini-gel glass plates and assemble on gel pouring rig. Add 8 μL of TEMED and 90 μL of 10% APS to gel

solution, quickly pipette gel solution between glass plates and place comb. Should take ~15 minutes to polymerize.

Running samples:

Pre-run gel in 1X TBE at 220 V for 15 minutes. Simultaneously heat gel samples for 15 minutes at 95°C.

After pre-running the gel, clear the gel wells using a blunt syringe and load samples. Run at 220 V for ~75 minutes.

In vitro* exonuclease assay:*10x Exonuclease Assay Buffer:**

TRIS 100 mM
 NaCl 100 mM
 MgCl₂*6H₂O 20 mM
 pH to 7.5, syringe filter

RNA Loading Buffer:

Formamide 9.5 mL
 Bromophenol blue
 EDTA (Final: 2 mM)

RNA Binding Buffer:

KCl 100 mM
 HEPES 100 mM
 MgCl₂*6H₂O 20 mM
 pH 7.4 and filter

Reagent	[Start]	[Final]
DEPC H2O		
10xBuffer	10X	1X
DTT	10 mM	1 mM
Nsp10	10 μM	500 nM
Nsp14	2.5 μM	125 nM
RNA (ss or ds)	5 μM	250 nM

Final rxn vol =

20 μL each.

Annealing RNA primer and template for dsRNA:

Want a 1:1.2 ratio of primer;template

Combine RNA with RNA binding buffer (1 μL per 20 μL total volume) and DEPC treated water.

Heat at 95°C for 5 min, turn heat block off and allow to cool in cooling heat block for 1 hr.

Move to RT for 15 min. Samples can now be stored at -20°C.

Rxn Protocol:

Dilute proteins in 1X reaction buffer

Combine reagents, incubate at 30°C for 15 minutes

To quench reaction, mix 5 μL of sample with 10 μL of RNA Loading Buffer.

Gel:

Combine 4.8 g of urea with 0.5 mL of 10X TBE, 5 mL of 30% acrylamide:bis-acrylamide 19:1, and DEPC treated water (to 10 mL) and dissolve the urea. Clean mini-gel glass plates and assemble on gel pouring rig. Add 8 μL of TEMED and 90 μL of 10% APS to gel solution, quickly pipette gel solution between glass plates and place comb. Should take ~15 minutes to polymerize.

Running samples:

Pre-run gel in 1X TBE at 220 V for 15 minutes. Simultaneously heat gel samples for 15 minutes at 95°C.

After pre-running the gel, clear the gel wells using a blunt syringe and load samples.

Run at 220 V for ~60 minutes.

In vitro* helicase assay:*10X Non-Denaturing Loading Buffer:**

10 mM TRIS
 1 mM EDTA
 50% Glycerol (v/v)
 Tiny! Bit of Bromophenol Blue
 2% SDS
 To 10 mL w/ DEPC H₂O

6% PAGE Gel

2 mL 30% Acrylimide:Bis-Acrylimide 19:1
 0.5 mL 10X TBE Running Buffer
 To 10 mL w/ DEPC H₂O
 90 µL of 10% APS
 8 µL of TEMED

Reagent	[Start]	[Final]
MilliQ H ₂ O		
10X Buffer	10X	1X
DTT	10 mM	1 mM
Single-Stranded DNA	5 µM	250 nM
Duplex DNA	5 µM	250 nM
Trap	25 µM	1.25 µM
ATP	10 mM	0.5 mM
Helicase	10 µM	500 nM

Reaction volume: 20 µL

DNA Substrates:

- **Single-Stranded DNA:** dilute to 5 µM
- **Trap:** dilute to 5 µM (For details on the use of the trap DNA, check “**Appendix 2**” of this thesis.
- **Double-Stranded DNA (Duplex, 5 µM):** Mix reagents, heat at 95 °C for five minutes, let cool in heat block for 1 hour, set at room temperature for 15 min then move to ice. Can be stored at -20°C.

Reaction:

- 1.) Combine reagents and then gently mix.
 - a. Note: Include a reaction of just ssDNA Probe to track ssRNA on the gel
- 2.) Set all reactions at 30°C for 15 minutes.
- 3.) After 15 minutes, move to ice.
- 4.) Combine 9 µL of reaction mixture with 1 µL of 10X Loading buffer.
- 5.) Load samples onto gel, then run gel for 45 minutes at 70V
 - a. Pre chill the gel and gel running rig on ice for ~1 hr before running, and then run on ice.
- 6.) Image gel with Typhoon 9200 using Cy2 excitation/emission

Electrophoretic mobility shift assay:

This example protocol is testing formation of the proofreading RTC. Other viral enzymes and cofactors can be added to test different and/or larger RTC formation.

10X Non-Denaturing Loading Buffer:

10mM TRIS
1mM EDTA
50% Glycerol (v/v)
Bromophenol Blue
To 10 mL w/ DEPC H₂O

4.5% PAGE Gel:

1.5 mL 30% Acrylamide:Bis-Acrylamide
19:1
1 mL 10X TBE Running Buffer
To 10 mL w/ DEPC H₂O
90 µL 10% APS
8 µL TEMED

RNA annealing buffer (10 mL):

HEPES 100 mM
KCl 100 mM
MgCl₂ 20 mM
pH to 7.4
Filter with 0.45 µm filter

10X HEPES (50 mL):

HEPES 250 mM
NaCl 500 mM
MgCl₂ 20 mM
pH to 8 with HCl, syringe filter
aliquots stored at -20°C

NOTE: Several buffers can work for EMSAS depending on the CoV RTC being studied. This includes using pH 7.5 (HEPES) or pH 8 (TRIS) buffering agents with added NaCl, KCl, Na-Glutamate, K-Glutamate (100-100 mM in 10X stock) and MgCl₂ (20 mM in 10X stock).

Reagent	[Start]	[Final]
DEPC H ₂ O		
10x Buffer	10x	1x
DTT	10 mM	1 mM
Nsp 7	20 µM	1 µM
Nsp 8	30 µM	1.5 µM
Nsp 12	10 µM	0.5 µM
Primer:Temp	12 µM	0.6 µM
Nsp10	45 µM	2.25 µM
Nsp14	30 µM	1.5 µM

Final rxn volume: 20 µL

Annealing RNA primer and template:

Want a 1:1.2 ratio of primer;template

Combine RNA with RNA binding buffer (1 µL per 20 µL total volume) and DEPC treated water.

Heat at 95°C for 5 min, turn heat block off and allow to cool in cooling heat block for 1 hr.

Move to RT for 15 min. Samples can now be stored at -20°C.

Gel Sample Prep:

- Dilute proteins in 1X reaction buffer

- Combine water, buffer, DTT and nsp's7+8+12 incubate at 25°C for 15 min
- Add RNA, incubate at 25°C for 15 min
- Add nsp10+14, incubate at 25°C for 15 min
- Remove 9 μL of each reaction and add to separate tube with 1 μL of 10x Non-Den loading buffer, run 4 μL of samples on a gel

Run Gel:

Run at 70 volts (10V cm^{-1} lengthwise) run in 1x TBE, run at 4°C (or on ice) with each component having been prechilled for about 1 – 1.25 hr. Run for 100 minutes.

Appendix 4: PDFs of coauthored publications

Interfering with nucleotide excision by the coronavirus 3'-to-5' exoribonuclease

Rukesh Chinthapatla¹, Mohamad Sotoudegan¹, Pankaj Srivastava¹, Thomas K. Anderson², Ibrahim M. Moustafa³, Kellan T. Passow⁴, Samantha A. Kennelly⁴, Ramkumar Moorthy⁴, David Dulin^{5,6}, Joy Y. Feng⁷, Daniel A. Harki⁴, Robert N. Kirchdoerfer², Craig E. Cameron^{1,*} and Jamie J. Arnold^{1,*}

¹Department of Microbiology and Immunology, The University of North Carolina at Chapel Hill School of Medicine, Chapel Hill, NC 27599, USA, ²Department of Biochemistry and Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI 53706, USA, ³Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA, ⁴Department of Medicinal Chemistry, University of Minnesota, Minneapolis, MN 55455, USA, ⁵Department of Physics and Astronomy, and LaserLAB Amsterdam, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands, ⁶Junior Research Group 2, Interdisciplinary Center for Clinical Research, Friedrich-Alexander-University Erlangen-Nürnberg (FAU), Cauerstr. 3, 91058 Erlangen, Germany and ⁷Gilead Sciences, Inc, Foster City, CA 94404, USA

Received August 11, 2022; Revised November 11, 2022; Editorial Decision November 17, 2022; Accepted November 27, 2022

ABSTRACT

Some of the most efficacious antiviral therapeutics are ribonucleos(t)ide analogs. The presence of a 3'-to-5' proofreading exoribonuclease (ExoN) in coronaviruses diminishes the potency of many ribonucleotide analogs. The ability to interfere with ExoN activity will create new possibilities for control of SARS-CoV-2 infection. ExoN is formed by a 1:1 complex of nsp14 and nsp10 proteins. We have purified and characterized ExoN using a robust, quantitative system that reveals determinants of specificity and efficiency of hydrolysis. Double-stranded RNA is preferred over single-stranded RNA. Nucleotide excision is distributive, with only one or two nucleotides hydrolyzed in a single binding event. The composition of the terminal basepair modulates excision. A stalled SARS-CoV-2 replicase in complex with either correctly or incorrectly terminated products prevents excision, suggesting that a mispaired end is insufficient to displace the replicase. Finally, we have discovered several modifications to the 3'-RNA terminus that interfere with or block ExoN-catalyzed excision. While a 3'-OH facilitates hydrolysis of a nucleotide with a normal ribose configuration, this substituent is not required for a nucleotide with a planar ribose configuration such as that present in the antiviral nucleotide produced by viperin. Design of

ExoN-resistant, antiviral ribonucleotides should be feasible.

INTRODUCTION

The emergence of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the causative agent of COVID-19, has led to a global pandemic and caused immeasurable consequences to humankind even more substantial than the incidence of disease and death. While the development of safe and effective vaccines has diminished overall morbidity and mortality, transmission of SARS-CoV-2 continues. Current therapeutics against SARS-CoV-2 infection include polymerase inhibitors: remdesivir and molnupiravir (1,2); and a protease inhibitor: nirmatrelvir, which is boosted with ritonavir (3). All of these therapeutic agents have complications that are tolerable in the midst of the pandemic. However, safer and more effective antiviral agents are needed against multiple targets to support the use of drug cocktails to maximize therapeutic efficacy and minimize the possibility of resistance.

Ribonucleos(t)ide analogs are among the most effective antiviral therapeutics for treatment of RNA virus infections (4–7). This class of compounds is generally administered as the ribonucleoside or ribonucleoside monophosphate pro-drug. Cellular kinases then produce the active metabolite, a triphosphorylated ribonucleotide (rNTP) (7). Utilization of the rNTP analog by viral polymerases leads to one or more consequences: chain termination, lethal mutagenesis,

*To whom correspondence should be addressed. Tel: +1 919 445 2922; Fax: +1 919 962 8103; Email: jamie_arnold@med.unc.edu
Correspondence may also be addressed to Craig E. Cameron. Tel: +1 919 966 9699; Fax: +1 919 962 8103; Email: craig.cameron@med.unc.edu

backtracking, pausing, and/or recombination, all of which exhibit an antiviral effect (5,6,8–10).

The coronavirus genome is on the order of 30 000 nt, among the largest RNA genome known (11). Because the fidelity of viral RNA polymerases is usually lower than 1 error per 10 000 nt, how coronaviruses resist lethal mutagenesis has been a longstanding question for those studying these viruses (12–14). Efforts initiated in response to emergence of the first SARS-CoV led to the discovery of a 3'-to-5' proofreading exoribonuclease, termed ExoN (15). ExoN is a member of the DEDDh/DEEDh subfamily in the DEED family of exonucleases (13,14,16). These enzymes use a two-metal-ion mechanism for catalysis (13,14,17,18). ExoN is composed of a 1:1 complex of non-structural proteins 14 (60 kDa) and 10 (15 kDa). Nsp14 harbors two catalytic domains: an N-terminal exoribonuclease and a C-terminal methyltransferase (13,14). Nsp10 serves as an accessory factor, stabilizing the active conformation of the exoribonuclease active site and thereby stimulating its activity (19–22).

Consistent with a role of ExoN in proofreading and therefore genome stability, previous studies have shown that genetic inactivation is most often lethal (15,23,24). For those coronaviruses, which are viable in the absence of active ExoN, the genomes contain a higher mutational load. Also, these viruses exhibit enhanced sensitivity to some antiviral ribonucleotides (25). Collectively, these observations suggest that inhibitors of ExoN will exhibit antiviral activity and may synergize with ribonucleotide-based inhibitors of the viral polymerase (26). Moreover, an understanding of the determinants of the substrate nucleotide that promote or interfere with excision may guide the design of ExoN-resistant ribonucleotide analogs.

Proofreading DNA exonucleases are generally a subunit of the DNA polymerase holoenzyme and exhibit a preference for mispaired ends (27). This circumstance appears to reflect the preferential partitioning of the mispaired end to the active site of the exonuclease instead of the active site of the polymerase (27,28). How proofreading by ExoN is initiated is not known. Given the dimensions of ExoN and the location of the polymerase active site within the SARS-CoV-2 replication-transcription complex (RTC), an active-site-switching mechanism is unlikely (29,30). It is easy to imagine how an end that cannot be extended by the RTC could become a substrate for repair by ExoN after the RTC dissociates. The ability for nucleotide analogs like remdesivir and molnupiravir to display efficacy in the presence of active ExoN may reflect the inability of these analogs to perturb elongation upon incorporation, with the antiviral activity manifested at the level of the analog-substituted template (31–33).

We have expressed and purified ExoN, and established a robust, quantitative system to study the determinants of the scissile phosphodiester bond and terminal ribonucleotide driving efficient excision by ExoN. ExoN prefers a primed-template-like dsRNA substrate over a ssRNA substrate and cannot access the 3'-terminus at a nick. ExoN only cleaves one or two nucleotides in a single binding event, as expected for a proofreading enzyme (27,28). Cleavage can be completely inhibited by the presence of the phosphorothioate Rp isomer at the scissile phosphodiester bond. A mispaired

end does not appear to be highly favored by ExoN relative to a paired end in the absence or presence of a replicating RTC. However, stalling the RTC at the 3'-end of nascent RNA blocks excision regardless of the nature of the base-pair. The inability of a mispaired end to promote dissociation of the RTC suggests a role for additional factors in this process. Finally, we identify modifications to the 3'-terminal nucleotide that interfere with excision by ExoN. The most unexpected finding was that the ribose conformation determines whether the 3'-OH is required for efficient turnover. The antiviral nucleotide, ddhCTP, produced by the antiviral protein, viperin, is readily excised by ExoN. This molecule lacks a 3'-OH but also lacks sugar pucker because of the presence of a double bond forcing the ribose into a planar conformation. This observation is consistent with repair of ddhC-terminated RNA as a major driver for acquisition of ExoN and evolution of its substrate specificity.

MATERIALS AND METHODS

Materials

DNA oligonucleotides and dsDNA fragments, GBLOCKS, were from Integrated DNA Technologies. RNA oligonucleotides were either from Horizon Discovery Ltd. (Dharmacon) or Integrated DNA Technologies. Restriction enzymes and T4 PNK (3'-phosphatase minus) were from New England Biolabs. IN-FUSION HD enzyme was from TakaraBio. Phusion DNA polymerase and T4 polynucleotide kinase were from ThermoFisher. pBirACm plasmid DNA was from Avidity. Streptactin XT 4F High-Capacity resin was from IBA Life Sciences. [γ - 32 P]ATP (6000 Ci/mmol) and [α - 32 P]ATP (3000 Ci/mmol) were from Perkin Elmer. Nucleoside 5'-triphosphates (ultrapure solutions) were from Cytiva. pET16b-RtcA-NTerm-His expression plasmid was provided by Stewart Shuman (Sloan Kettering) (34). Cytidine 5'-O-(1-thiotriphosphate) and 3'-deoxycytidine 5'-triphosphate were from TriLink. Adenosine 5'-O-(1-thiotriphosphate) (Sp isomer) was from BioLog. Remdesivir-terminated RNA was synthesized by both Dharmacon and the Harki lab as previously described (35). Remdesivir was provided to Dharmacon by Gilead Sciences, and this was chemically converted to the phosphoramidite to be synthetically incorporated into RNA. ddhCTP was synthesized by the Harki lab as previously described (36); 2'-C-methylcytidine 5'-triphosphate was provided by Gilead Sciences. All other reagents were of the highest grade available from MilliporeSigma, VWR, or Fisher Scientific.

Construction of modified pSUMO vectors containing AviTag

The pSUMO system allows for production of SUMO fusion proteins containing an amino-terminal affinity tag fused to SUMO that can be purified by affinity chromatography and subsequently processed by the SUMO protease, Ulp1 (37). After cleavage this will produce an authentic untagged protein target of interest. The pSUMO vector (LifeSensors) (37) was modified such that the coding sequence for the six-histidine tag was replaced with a DNA sequence coding for an AviTag codon optimized for bacterial expression (38,39). The AviTag is a short 15 amino acid se-

quence (GLNDIFEAQKIEWHE) that can be specifically biotinylated on the lysine residue by the biotin ligase, BirA (38,39). The construct contains two tags separated by a short linker to increase the affinity to the chromatography resin during purification. The biotinylated AviTag allows affinity purification of the fusion protein to be isolated using streptavidin or Strep-Tactin resins. The DNA sequences (GBlocks encoding the tags) were cloned into pSUMO using XbaI and SalI by IN-FUSION. The final constructs (pAviTag_SUMO) were confirmed by sanger sequencing performed by Genewiz.

Codon-optimized sequence for AviTag

5'ATGGGACTAAATGATATATTTGAAGCTCAAAA
GATCGAGTGGCACGAGGGTGGTGGCAGCGGTG
GCGGCTCCGGCGGTAGCGGCCTGAACGACATC
TTCGAGGCGCAGAAAATTGAATGGCATGAAGG
TGGCTCTAGCGGTGGT3'

(amino acid sequence: MGLNDIFEAQKIEWHEGG
GSGGGSGSGLNDIFEAQKIEWHEGGSSGG)

Construction of SARS-CoV-2 nsp10 and nsp14 bacterial expression plasmids

The SARS-CoV-2 nsp10 and nsp14 genes were codon optimized for expression in *E. coli* and obtained from Genescript. The amino acid sequences for nsp10 and nsp14 were derived from SARS-CoV-2 isolate 2019-nCoV/USA-WA1/2020 (GenBank MN985325.1). The genes were amplified by PCR using Phusion DNA polymerase. The nsp10 gene was amplified using the synthetic nsp10 gene as template and DNA oligonucleotides (5'-GAACAGATTGGAGGTGCCGGGAATGCTACGGAA-3' and 5'-CCGCAAGCTTGTGACTTATCATTGAAGCATA GGTTACGCAA-3'). The nsp14 gene was amplified using the synthetic nsp14 gene as template and DNA oligonucleotides (5'-GAACAGATTGGAGGTGCCGA GAACGTTACAGGT and 5'-CCGCAAGCTTGTGCGAC TTATCATTGAAGGCCAGTAAACGTATTCCA-3').

The PCR products were gel purified and cloned into either the pAviTag-pSUMO bacterial expression plasmids using BsaI and SalI. The final constructs were confirmed by sanger sequencing performed by Genewiz. To construct a catalytically inactive nsp14, the WT nsp14 expression plasmid was modified such that D90 and E92 were both changed to alanine. This was performed using Quickchange mutagenesis using the WT nsp14 expression plasmid as template and DNA oligonucleotides (5'-GCGTGGATTG GTTTTGCTGTTGCGGGTTGCCACGCGACCCCGT-3' and 5'-ACGGGTGCGGTGGCAACCCGCAAC AGCAAACCAATCCACGC-3'). The final construct was confirmed by sanger sequencing performed by Genewiz.

Codon optimized sequence for SARS-CoV-2 nsp10

5'GCCGGGAATGCTACGGAAAGTTCCAGCTAACT
GACCGTTCTTAGCTTTTGTGCTTTTGCAGTCGAT
GCAGCGAAAGCGTATAAGGACTATCTGGCGTC
AGGGGGACAACCCATTACTAACTGTGTCAAGA

TGCTGTGTACCCATACCGGCACGGGTCAAGCG
ATTACTGTTACACCAGAAGCTAACATGGACCA
GGAATCTTTTGGTGGTGGCCAGTTGCTGCTTGT
ACTGCCGCTGTCATATCGATCACCCCAATCCAAA
AGTTTTCTGCGATCTGAAGGGAAAATACGTGC
AAATCCCCACCACTTGTGCTAATGACCCGGTC
GGATTTACGCTGAAGAACACCGTTTGTACTGTTT
GCGGGATGTGGAAAGGGTATGGGTGTTCTTGC
GACCAGTTGCGTGAACCTATGCTTCAA3'.

Codon optimized sequence for SARS-CoV-2 nsp14

5'GCGGAGAACGTTACAGGTTTATTTAAGGATTG
CTCTAAAGTAATTACCGGCCTGCACCCCAACGC
AGGCACCAACTCATCTTAGCGTGGATACAAA
TTTAAGACAGAAGGACTGTGTGTGGACATTCC
TGGCATCCCAAAGGACATGACATACCGCCGTT
TGATCTCCATGATGGGGTTCAAAAATGAACTAC
CAGGTAAACGGATACCCTAATATGTTTCATTAC
ACGTGAGGAGGCGATTTCGTCATGTCGCGCCT
GGATCGGATTCGACGTAGAAGGTTGCCACGCC
ACCCGTGAGGCTGTGGGGACGAACTTACCCCT
TCAGCTTGGCTTCTCAACTGGGGTAAACTTGG
TGGCCGTCCCGACAGGGTATGTTGACACTCCT
AATAACACTGATTTCTCGCGTGTATCTGCAAAGC
CACCACAGGGGACCAGTTCAAACACACTGATC
CCCCTGATGTATAAGGGTCTTCCCTTGGAAATG
GGTCCGTATTAATAATCGTCCAGATGCTGTGAG
ACACCCTTAAGAATCTGTCAGATCGTGTGGTA
TTTGTATTGTGGGCGCACGGATTCGAGTTAAC
AAGCATGAAATATTTTGTGAAAATTGGCCCCG
AACGCACATGCTGCTTATGCGATCGTCGCGCT
ACTTGCTTTAGTACTGCTTCAGACACTTATGCCT
GCTGGCACCCTCTATTGGATTTGACTACGTGTA
TAACCCATTTCATGATTGATGTCCAGCAGTGGG
GCTTCACCGGGAACCTTGCAGTCCAACCATGAC
CTTTATTGTCAGGTTACCGGAAATGCCACCGT
GGCAACTGCGACGCGATTATGACACGCTGTC
TGGCGGTACATGAGTGCTTTGTAAAGCGTGTG
GATTGGACCATCGAGTATCCAATCATTGGAGA
CGAACTTAAGATCAATGCCGCATGCCGTAAAG
TTCAACACATGGTAGTAAAGGCCGCCCTTCTT
GCGGATAAGTTTCCGGTTCTGCATGACATTGG
CAACCCTAAGGCGATTAAGTGTGTCCCGCAGG
CGGATGTGCAATGGAAATTCTATGACGCGCAA
CCCTGCTCGGATAAAGCATATAAAATCGAAGA
GCTGTTTTATTCATACGCTACGCATTCCGACAAG
TTTACAGATGGCGTTTGTCTTTTTTGGAAATTGTA
ACGTTGATCGCTACCCGGCGAACTCAATCGTT
TGCCGCTTTGACACACGCTGTGCTGTCTAACTGA
ACTTGCCTGGTTGCGATGGAGGCTCGTTGTAT
GTTAATAAACATGCGTTTCATACCCCGCCTTCG
ACAAGTCCGCTTTTCGTAAACCTGAAGCAGTTG
CCATTTTTCTACTATAGCGACTACCCGTGCGAGT
CCCACGGTAAGCAAGTAGTGTCTGACATTGAT
TATGTACCTTTAAAAAGTGCTACCTGCATCACCC
GTTGCAACTTGGGCGGAGCGGTTTGGCGCCAC
CATGCGAACGAATATCGCTTATACCTTGATGCCT
ATAATATGATGATTAGCGCGGGATTTAGCCTT
TGGGTTTATAAACAGTTCGATACTTATAACCTGT
GGAATACGTTTACTCGCCTTCAA3'.

Expression and purification of SARS-CoV-2 nsp10

Escherichia coli BL21(DE3)pBirACm competent cells were transformed with the pAviTag.SUMO-SARS-CoV-2-nsp10 plasmid for protein expression. BL21(DE3)pBirACm cells containing the pAviTag.SUMO-SARS-CoV-2-nsp10 plasmid were grown in 100 mL of media (NZCYM) supplemented with kanamycin (K25, 25 µg/ml) and chloramphenicol (C20, 20 µg/ml) at 37°C until an OD₆₀₀ of 1.0 was reached. This culture was then used to inoculate 4 l of K25,C20 media to an OD₆₀₀ = 0.1. Biotin (25 mM in 500 mM Bicine pH 8.0) was added to the media to a final concentration of 50 µM. The cells were grown at 37°C to an OD₆₀₀ of 0.8 to 1.0, cooled to 25°C and then IPTG (500 µM) was added to induce protein expression. Cultures were then grown for an additional 4 h at 25°C. Cells were harvested by centrifugation (6000 × g, 10 min) and the cell pellet was washed once in 200 ml of TE buffer (10 mM Tris, 1 mM EDTA), centrifuged again, and the cell paste weighed. The cells were then frozen and stored at -80°C until used. Frozen cell pellets were thawed on ice and suspended in lysis buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol, 1.4 µg/ml leupeptin, 1.0 µg/ml pepstatin A and two Roche EDTA-free protease tablet per 5 g cell pellet), with 5 ml of lysis buffer per 1 gram of cells. The cell suspension was lysed by passing through a French press (SLM-AMINCO) at 15 000 psi. After lysis, phenylmethylsulfonylfluoride (PMSF) and NP-40 were added to a final concentration of 1 mM and 0.1% (v/v), respectively. While stirring the lysate, polyethylenimine (PEI) was slowly added to a final concentration of 0.25% (v/v) to precipitate nucleic acids from cell extracts. The lysate was stirred for an additional 30 min at 4°C after the last addition of PEI, and then centrifuged at 75 000 × g for 30 min at 4°C. The PEI supernatant was then loaded onto a Strep-Tactin XT 4F HC resin (IBA Life Sciences) at a flow rate of 1 ml/min (approximately 1 mL bed volume per 100 mg total protein) equilibrated with buffer A (25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol). After loading, the column was washed with twenty column volumes of buffer A. The resin was then suspended in two column volumes of buffer A and Ulp1 (5 µg per 1 ml bed volume) was added with the resin overnight at 4°C to cleave the SUMO-nsp10 fusion protein. The column was then washed with 5 column volumes of buffer A and fractions were collected and assayed for purity by SDS-PAGE. The protein concentration was determined by measuring the absorbance at 280 nm by using a Nanodrop spectrophotometer and using a calculated molar extinction coefficient of 13 700 M⁻¹ cm⁻¹. Purified, concentrated protein was aliquoted and frozen at -80°C until use. Typical nsp10 yields were 1 mg/1 g of *E. coli* cells.

Expression and purification of SARS-CoV-2 nsp14

Escherichia coli BL21(DE3)pBirACm competent cells were transformed with the pAviTag.SUMO-SARS-CoV-2-nsp14 plasmid for protein expression. BL21(DE3)pBirACm cells containing the pAviTag.SUMO-SARS-CoV-2-nsp14 plasmid were grown in 100 ml of media (NZCYM) supplemented with kanamycin (K25, 25 µg/ml) and chloram-

phenicol (C20, 20 µg/ml) at 37°C until an OD₆₀₀ of 1.0 was reached. This culture was then used to inoculate 4 l of K25,C20 media to an OD₆₀₀ = 0.1. Biotin (25 mM in 500 mM Bicine pH 8.0) was added to the media to a final concentration of 50 µM. The cells were grown at 37°C to an OD₆₀₀ of 0.8 to 1.0, cooled to 25°C and then IPTG (500 µM) was added to induce protein expression. Cultures were then grown for an additional 4 h at 25°C. Cells were harvested by centrifugation (6000 × g, 10 min) and the cell pellet was washed once in 200 mL of TE buffer (10 mM Tris, 1 mM EDTA), centrifuged again, and the cell paste weighed. The cells were then frozen and stored at -80°C until used. Frozen cell pellets were thawed on ice and suspended in lysis buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol, 1.4 µg/ml leupeptin, 1.0 µg/ml pepstatin A and two Roche EDTA-free protease tablet per 5 g cell pellet), with 5 ml of lysis buffer per 1 g of cells. The cell suspension was lysed by passing through a French press (SLM-AMINCO) at 15 000 psi. After lysis, phenylmethylsulfonylfluoride (PMSF) and NP-40 were added to a final concentration of 1 mM and 0.1% (v/v), respectively. While stirring the lysate, polyethylenimine (PEI) was slowly added to a final concentration of 0.25% (v/v) to precipitate nucleic acids from cell extracts. The lysate was stirred for an additional 30 min at 4°C after the last addition of PEI, and then centrifuged at 75 000 × g for 30 min at 4°C. The PEI supernatant was then loaded onto a Strep-Tactin XT 4F HC resin (IBA Life Sciences) at a flow rate of 1 ml/min (~1 ml bed volume per 100 mg total protein) equilibrated with buffer A (25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol). After loading, the column was washed with twenty column volumes of buffer A. The resin was then suspended in two column volumes of buffer A and Ulp1 (5 µg per 1 ml bed volume) was added with the resin overnight at 4°C to cleave the SUMO-nsp14 fusion protein. The column was then washed with 5 column volumes of buffer A and fractions were collected and assayed for purity by SDS-PAGE. The protein concentration was determined by measuring the absorbance at 280 nm by using a Nanodrop spectrophotometer and using a calculated molar extinction coefficient of 93 625 M⁻¹ cm⁻¹. Purified, concentrated protein was aliquoted and frozen at -80°C until use. Typical nsp14 yields were 0.2 mg/1 g of *E. coli* cells. The catalytically inactive nsp14 D90A E92A was expressed and purified using the exact same conditions as WT nsp14.

SARS CoV-2 nsp7, nsp8 and nsp12 recombinant proteins

Expression and purification of SARS CoV-2 nsp7, nsp8 and nsp12 are described in detail in (8).

Expression and purification of RtcA

Escherichia coli Rosetta(DE3) competent cells were transformed with the pET16b-RtcA-NTerm-His plasmid for protein expression. *E. coli* Rosetta(DE3) cells containing the pET16b-RtcA-NTerm-His plasmid were grown in 100 ml of A300,C60-supplemented ZYP-5052 auto-induction media at 37°C (40,41). The cells were grown at 37°C to an

OD₆₀₀ of 0.8 to 1.0, cooled to 15°C and then grown for 36–44 h. After ~40 h at 15°C the OD₆₀₀ reached ~10–15. Cells were harvested by centrifugation (6000 × *g*, 10 min) and the cell pellet was washed once in 200 ml of TE buffer (10 mM Tris, 1 mM EDTA), centrifuged again, and the cell paste weighed. The cells were then frozen and stored at -80°C until used. Frozen cell pellets were thawed on ice and suspended in lysis buffer (25 mM HEPES pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol, 5 mM imidazole, 1.4 μg/ml leupeptin, 1.0 μg/ml pepstatin A and two Roche EDTA-free protease tablet per 5 g cell pellet), with 5 ml of lysis buffer per 1 g of cells. The cell suspension was lysed by passing through a French press (SLM-AMINCO) at 15 000 psi. After lysis, phenylmethylsulfonylfluoride (PMSF) and NP-40 were added to a final concentration of 1 mM and 0.1% (v/v), respectively. The lysate was centrifuged at 75 000 × *g* for 30 min at 4°C. The lysate was then loaded onto Qiagen Ni-Spin columns (a total of 5 mL per one Ni spin column). After loading, the column was washed with two 0.5 ml volumes of buffer B (25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM TCEP, 20% glycerol) with 5 mM imidazole, then washed with two 0.5 ml volumes of buffer B with 50 mM imidazole and then eluted in four 0.1 ml volumes of buffer B with 500 mM imidazole. Elution fractions were collected and assayed for purity by SDS-PAGE. The protein concentration was determined by measuring the absorbance at 280 nm by using a Nanodrop spectrophotometer and using a calculated molar extinction coefficient of 11 585 M⁻¹ cm⁻¹. Purified, concentrated protein was aliquoted and frozen at -80°C until use. RtcA yields were 0.5 mg/2 g of *E. coli* cells.

5'-³²P-labeling of RNA substrates

RNA oligonucleotides were end-labeled by using [γ-³²P]ATP and T4 polynucleotide kinase. Reaction mixtures, with a typical volume of 50 μl, contained 0.5 μM [γ-³²P]ATP, 10 μM RNA oligonucleotide, 1 × kinase buffer, and 0.4 unit/μl T4 polynucleotide kinase. Reaction mixtures were incubated at 37°C for 60 min and then held at 65°C for 5 min to heat inactivate T4 PNK. For RNAs containing a 3' phosphate T4 PNK (minus 3' phosphatase) was used using the same reaction conditions.

Cyclization reactions to produce 2',3'-cyclic phosphate containing RNAs

Reactions contained 25 mM HEPES pH 7.5, 10 mM MgCl₂, 10 mM TCEP, 50 mM NaCl, 100 μM ATP, 1 μM ³²P-labeled RNA (3'-phosphate termini) and 5 μM RtcA. Reactions were performed at 37°C for 30 min. Reactions were quenched by addition of EDTA to a final concentration of 10 mM and placed at 65°C to heat inactivate RtcA enzyme.

Annealing of dsRNA substrates

dsRNA substrates were produced by annealing 10 μM RNA oligonucleotides in T₁₀E₁ [10 mM Tris pH 8.0 and 1 mM EDTA] and 50 mM NaCl in a Progene Thermocycler (Techne). Annealing reaction mixtures were heated to

90°C for 1 min and slowly cooled (5°C/min) to 10°C. Specific scaffolds are described in the figure legends.

Native PAGE

5'-P³²-labeled ssRNA and dsRNA substrates were mixed with an equal volume of loading buffer (10% glycerol, 0.025% bromophenol blue and 0.025% xylene cyanol) and loaded onto a 10% or 20% acrylamide, 0.5% bisacrylamide native polyacrylamide gel containing 1 × TBE (89 mM Tris base, 89 mM boric acid and 2 mM EDTA). Electrophoresis was performed in 1 × TBE at 15 mA. Gels were visualized by using a Phosphorimager.

SARS-CoV-2 nsp10/nsp14-catalyzed exoribonuclease assays

Reactions contained 25 mM HEPES pH 7.5, 2 mM MgCl₂, 1 mM TCEP, 10 mM KCl, and 50 mM NaCl and were performed at 30°C. Reactions were quenched by addition of EDTA to a final concentration of 25 mM. Specific concentrations of RNA substrate, ExoN, along with any deviations from the above, are indicated in the appropriate figure legend. Typical concentrations for RNA substrate and enzyme were between 0.01 to 2 μM. Enzymes were diluted immediately prior to use in 25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM TCEP, and 20% glycerol. SARS-CoV-2 nsp10 was pre-mixed with SARS-CoV-2 nsp14 on ice in 25 mM HEPES, pH 7.5, 500 mM NaCl, 2 mM TCEP, and 20% glycerol for 5 min prior to initiating the reaction with ExoN. The volume of enzyme added to any reaction was always less than or equal to one-tenth the total volume. The ExoN concentration refers to the nsp14 concentration and the ratio of nsp14 to nsp10 is also indicated in cases where nsp10 was in excess of nsp14. For example, 0.1 μM ExoN (1:5) refers to final concentrations of nsp14 of 0.1 μM and nsp10 at 0.5 μM. Products were resolved from substrates by denaturing PAGE.

Incorporation of modified nucleoside triphosphates into dsRNA substrates prior to challenge for removal by ExoN

Reactions contained 25 mM HEPES pH 7.5, 5 mM MgCl₂, 1 mM TCEP and 50 mM NaCl and were performed at 30°C. Human mitochondrial RNA polymerase, POLRMT, was used to incorporate modified nucleoside triphosphates. POLRMT was expressed and purified as described previously (42). 1 μM POLRMT was mixed with 1 μM ³²P-labeled dsRNA nucleic acid scaffold (primed-template) in the presence of 10 μM ATP, 10 μM UTP and either 10 μM CTP or 100 μM of the modified nucleoside 5'-triphosphate. The RNA primer was P9 (5'-CCGGGCGGC-3') and RNA template was T21 (Table 1). This primer-template pair allows ATP to be incorporated at the *n* + 1 position, UTP at *n* + 2 and the modified nucleoside 5'-triphosphate to be incorporated at the *n* + 3 position. Reactions were allowed to proceed for 30 min to allow complete extension to *n* + 3 at which point 0.1 μM ExoN (1:5) was added to the reaction. Reactions were quenched at various times by addition of 25 mM EDTA. Products were resolved from substrates by denaturing PAGE.

Table 1. RNA oligonucleotides. RNA substrates used in this study. The name, length, type of modification (if any) and sequence are indicated

#	Name	Length (nt)	Modification	Sequence
1	P10A	10	-	5'-CCGGGCGGCA-3'
2	P10C	10	-	5'-CCGGGCGGCC-3'
3	P10G	10	-	5'-CCGGGCGGCG-3'
4	P10U	10	-	5'-CCGGGCGGCU-3'
5	P10R	10	remdesivir	5'-CCGGGCGGCR-3'
6	P10A-2'd	10	2'-deoxy	5'-CCGGGCGGCA/2'd/-3'
7	P10A-3'd	10	3'-deoxy	5'-CCGGGCGGCA/3'd/-3'
8	P10A-3'P	10	3'-phosphate	5'-CCGGGCGGCA/3'P/-3'
9	Ps ₁ 10A	10	phosphorothioate	5'-CCGGGCGGC/s/A-3'
10	Ps ₂ 10A	10	phosphorothioate	5'-CCGGGCGG/s/CA-3'
11	Ps ₃ 10A	10	phosphorothioate	5'-CCGGGCGC/s/GCA-3'
12	Ps _{2,3} 10A	10	phosphorothioate	5'-CCGGGCG/s/G/s/CA-3'
13	P10U-2'F	10	2'-fluoro	5'-CCGGGCGGCU/2'F/-3'
14	P10U-2'NH ₂	10	2'-amino	5'-CCGGGCGGCU/2'NH ₂ /-3'
15	P10A-2'OMe	10	2'-O-methyl	5'-CCGGGCGGCA/2'OMe/-3'
16	P15	15	-	5'-AAGAAAGGAGGGAGG-3'
17	P25	25	-	5'-GGAAAGGGAAAGGGAAGGAGGAAGA-3'
18	P40	40	-	5'-GGAAAGGGAAAGGGAAGGAGGAAGAAAGAAAGGA GGGAGG-3'
19	P50	50	-	5'-GGAAAGGGAAAGGGAAGGAGGAAGAAAGAAAGGA GGGAGGCCGGGCGGCA-3'
20	T21	21	-	5'-CCCCCGAUGCCGCCCGCCCC-3'
21	T57	57	-	5'-CCCCCGAUGCCGCCCGCCUCCCUCCUUUCUUUCU CCUCCUUCCUUUCCCUUCC-3'
22	P9	9	-	5'-CCGGGCGGC-3'
23	P8	8	-	5'-CCGGGCGG-3'
24	P7	7	-	5'-CCGGGCG-3'
25	P6	6	-	5'-CCGGGC-3'
26	P5	5	-	5'-CCGGG-3'
27	P4	4	-	5'-CCGG-3'
28	P3	3	-	5'-CCG-3'
29	P2	2	-	5'-CC-3'

Incorporation of correct and incorrect nucleotides by SARS-CoV-2 replicase into dsRNA substrates prior to challenge for removal by ExoN

Reactions contained 25 mM HEPES pH 7.5, 2 mM MgCl₂, 1 mM TCEP and 50 mM NaCl and were performed at 30°C. Reactions contained 0.1 μM nsp12, 0.3 μM nsp7, 0.3 μM nsp8, 0.1 μM ExoN (1:5), 0.1 μM dsRNA primed/template (**P9-P40;T57**), 1 μM ATP, 0.1 μCi/μl [³²P]ATP, 1 μM UTP, and either 10 μM CTP or 100 μM of the modified nucleoside 5'-triphosphate. Initially, nsp12/7/8 was incubated with dsRNA substrate and ATP in the absence or presence of UTP for 60 min to form n + 1 or n + 2 (with UTP) product, at which point either CTP, 3'-dCTP, ddhCTP, 2'-C-Me-CTP was added to promote further extension. After 2 min, ExoN was added, and the reaction was quenched at various times by the addition of EDTA to 25 mM.

Denaturing PAGE analysis of Exonuclease-catalyzed reaction products

An equal volume of loading buffer (85% formamide, 0.025% bromophenol blue and 0.025% xylene cyanol) was added to quenched reaction mixtures and heated to 90°C for 5 min prior to loading 5 μl on a denaturing either 15% or 23% polyacrylamide gel containing 1× TBE (89 mM Tris base, 89 mM boric acid, and 2 mM EDTA) and 7 M urea. For reactions that contained dsRNA substrates an excess (50-fold) of unlabeled RNA oligonucleotide (trap strand) that is the exact same sequence to the ³²P-labeled RNA oligonucleotide in the reaction was present in the loading

buffer to ensure complete separation and release of the ³²P-labeled RNA oligonucleotide prior to gel electrophoresis (43). This procedure allows efficient strand separation of ³²P-labeled RNA oligos that are in the presence of their RNA complements (43). Electrophoresis was performed in 1× TBE at 90 W. Gels were visualized by using a PhosphorImager (GE) and quantified by using ImageQuant TL software (GE).

Data analysis

All gels shown are representative, single experiments that have been performed at least three to four individual times to define the concentration or time range shown with similar results. In all cases, values for parameters measured during individual trials were within the limits of the error reported for the final experiments. Data were fit by either linear or nonlinear regression using the program GraphPad Prism v7.03 (GraphPad Software Inc.).

RESULTS

Expression and purification of SARS-CoV-2 ExoN: nsp10 and nsp14

Many proteins and enzymes encoded by SARS-CoV-2 contain zinc ions coordinated by side chains of histidine and/or cysteine residues, including the nsp10 and nsp14 subunits of the 3'-to-5' exoribonuclease complex (ExoN) (11,13,14,19). For the nsp12 gene-encoded RdRp, the natural ligand has

been proposed to be a four-iron, four-sulfur cluster (44). To avoid the potential for displacement of metal ions during protein purification that were incorporated during protein expression, we used the AviTag-SUMO system, which avoids metal-affinity chromatography (Supplementary Figure S1A) (38,39). The AviTag is a 15 amino acid sequence (GLNDIFEAQKIEWHE) that is biotinylated on the lysine residue by biotin ligase, BirA, co-expressed with the fusion protein (38,39). The biotinylated AviTag-SUMO fusion protein binds to streptavidin or streptactin resins (Supplementary Figure S1B) (38,39). The protein of interest is released from the resin by cleavage at the carboxyl terminus of SUMO by the Ulp1 protease (Supplementary Figure S1C–F) (39). Purified proteins used in this study are shown in Figure 1A. We also purified a catalytically inactive derivative of nsp14, referred to as MUT (Supplementary Figure S1F).

ExoN Prefers a dsRNA substrate containing a recessed 3'-end, a 'primed-template'

We used ssRNA and dsRNA substrates of various lengths in the assay (Figure 1B, C, Table 1, Supplementary Figure S2). We chose these substrates because they can also be used as primers (P) and templates (T) for the replication-transcription complex (RTC). Native PAGE confirmed these substrates to be either single or double stranded (Figure 1C). We use denaturing polyacrylamide gel electrophoresis followed by phosphorimaging to monitor the hydrolysis of the ³²P-labeled RNA strand (indicated by an asterisk) in the absence (ssRNA) or presence of an annealed RNA strand (dsRNA). The concentration of ExoN and nsp14:nsp10 stoichiometry used was selected to reflect conditions used by others (19,20,23,29,45–52), to facilitate a comparison of the results obtained here to those. Our studies demonstrated that ssRNA was not a good substrate by ExoN (*P10A and *T57 in Figure 1D and Supplementary Figure S3A–C). The 3'-end of a blunt-ended duplex was also a poor substrate for ExoN (P50A:*T57 in Figure 1D and Supplementary Figure S3D). dsRNA with a recessed 3'-end was the most active substrate (*P10A:T21 in Figure 1D and Supplementary Figure S3F). Comparable results were obtained with P50A and T57 RNAs (Supplementary Figure S3). Mutagenesis of the nsp14 ligands required for Mg²⁺-dependent catalysis (D90A, E92A) inactivated ExoN (compare MUT to WT in Figure 1D) (13–16). Neither nsp10 nor nsp14 alone exhibited exonuclease activity on ssRNA or dsRNA substrates (Supplementary Figure S3B,C). Therefore, none of the activity measured here derived from a contaminating ribonuclease.

Products of the nsp15 endonuclease are not substrates for ExoN

Coronaviruses also encode an uridylyte-specific, ssRNA or dsRNA endonuclease, referred to as NendoU and is encoded by the nsp15-coding sequence of the SARS-CoV-2 genome (53,54). While the activity on dsRNA has been suggested to clear dsRNA that would otherwise activate intrinsic antiviral defenses (53,55,56), the possibility existed that this enzyme could contribute to a post-transcriptional,

mismatch-repair mechanism. For example, nsp15 might also cleave the phosphodiester backbone at or near a mismatch. If ExoN could cleave at a nick or at a 2'–3' cyclic-phosphate or 3'-phosphate terminus produced by nsp15, then such a repair mechanism might exist.

To test this possibility, we assembled substrates in which the P50A RNA was fragmented into three segments: P-1, P-2 and P-3 in the 5'-to-3' direction, respectively (Figure 2A, B and Supplementary Figure S2A). We evaluated three conditions: I, where P-3 was labeled; II, where P-3 was omitted and P-2 was labeled; and III, where P-2 was labeled (Figure 2A). The labeled, 3'-terminal P-3 primer was cleaved efficiently (I in Figure 2B). Similarly, P-2 was cleaved efficiently when present as the 3'-terminal RNA (II in Figure 2B). However, P-2 was not cleaved when embedded between P-1 and P-3 (III in Figure 2B), suggesting that ExoN is incapable of initiating hydrolysis from the 3'-OH at a nick.

Cleavage by nsp15 leaves a 3'-end with a 2'-3'-cyclic phosphate (cycP) that can be hydrolyzed to a 3'-phosphate (3'-OPO₃) and a 5'-end with a hydroxyl (Figure 2D). Using approaches developed by the Shuman laboratory (Supplementary Figure S4) (34), we prepared RNAs to test as substrates for ExoN that contained either a cycP or a 3'-OPO₃ at the recessed end of the dsRNA (Figure 2D, E). Neither of these RNAs served as substrates (Figure 2F); note that the presence of a 3'-phosphate and cycP group facilitates faster migration in the denaturing PAGE gel than the unmodified RNA with a 3'-OH (Figure 2F). So, even if the RNA downstream of the nick produced by nsp15 cleavage were removed, this product would be incapable of being degraded by ExoN without production of a 3'-hydroxyl.

Nucleotide hydrolysis by ExoN occurs in a distributive manner, as expected for a proofreading exonuclease.

Studies of nucleotide hydrolysis by ExoN that either employ conditions of ExoN in excess of substrate or evaluate product formation after long incubation times (30 min to hours) can give the appearance that ExoN is processive (18,23,29,45–48,50–52). In this case, processive means hydrolysis of multiples of ten nucleotides per binding event and is in contrast to a distributive enzyme, which would hydrolyze one or two nucleotides per binding event. Proofreading exonucleases associated with DNA polymerases have evolved to function in a distributive manner. This circumstance eliminates the mismatch without the need for re-synthesis of stretches of nucleic acid that were correctly basepaired, as would be the case if the exonuclease acted processively (27,28).

The ability to split the hydrolyzed strand into multiple components (Figure 2), offered the opportunity to create a substrate with sufficient dsRNA to form a processive, elongation complex using the SARS-CoV-2 RTC (30,57,58) but at the same time permit hydrolysis by ExoN to be monitored with single-nucleotide resolution. The substrate used is shown in Figure 3A. We evaluated utilization of this substrate by ExoN under two conditions. We refer to the first condition as substrate excess, which should reveal the length(s) of product formed by ExoN in a single binding event (left panel of Figure 3B). We refer to the second condition as enzyme excess (right panel of Figure 3B). Under

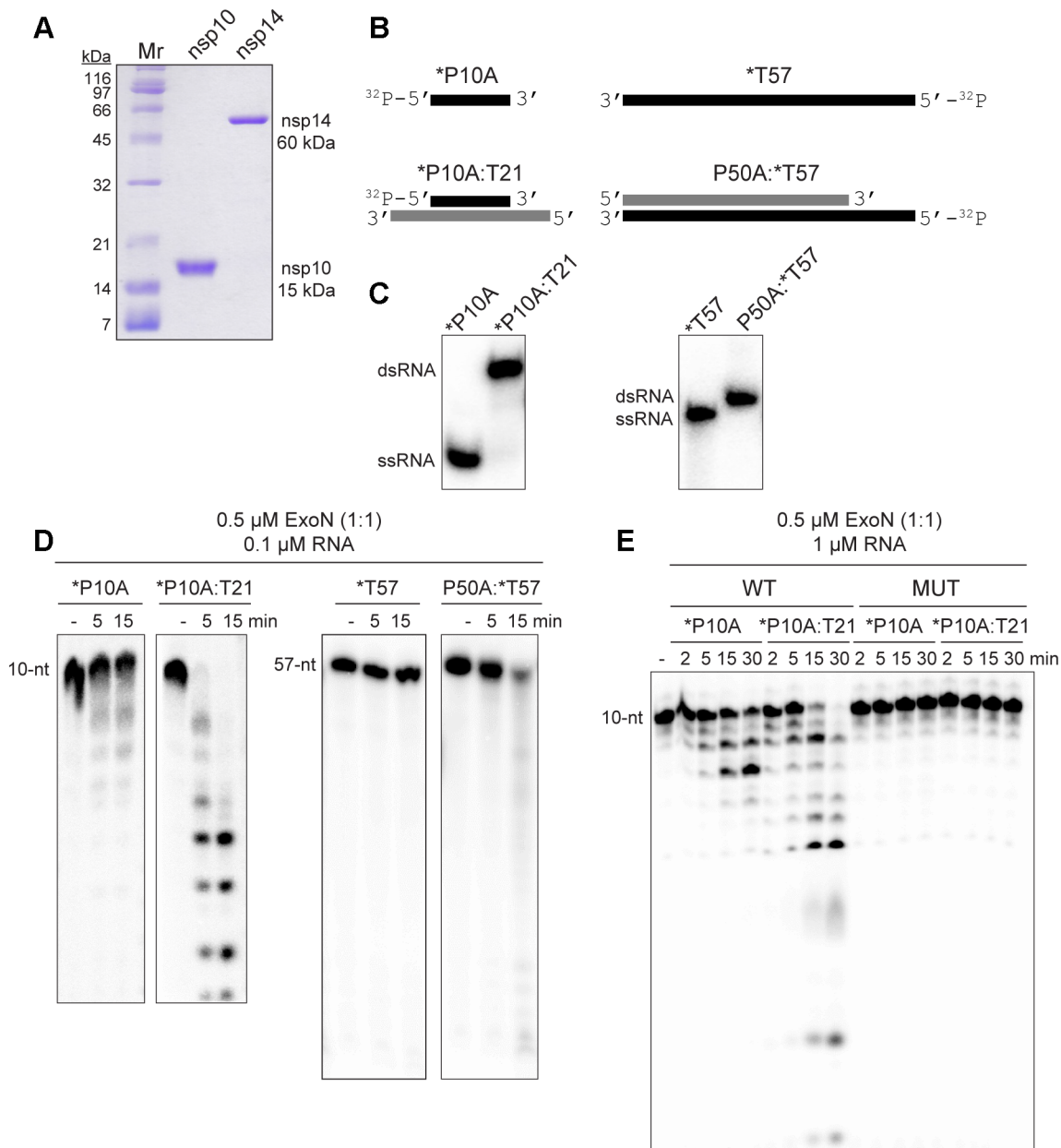


Figure 1. ExoN prefers primed-template-like dsRNA substrates. **(A)** Purified SARS-CoV-2 nsp14 (60 kDa) and nsp10 (15 kDa) proteins used in this study. Proteins (1 μg) were resolved on a 15% polyacrylamide gel containing SDS and stained with Coomassie. Broad-range molecular weight markers (Mr) and corresponding molecular weights are indicated. **(B)** Schematic of ssRNA and dsRNA substrates used to measure exoribonuclease activity. RNA sequences are provided in Table 1 and/or Supplementary Figure S2. RNAs were labeled on the 5'-end with ^{32}P , also indicated by an asterisk (*). RNAs are designated as primers (P) or templates (T). The numbers indicate the length of the RNA. When a base is designated, this reflects the 3'-terminal nucleotide. **(C)** ssRNA and dsRNA substrates evaluated by Native PAGE. **(D)** Evaluation of ExoN-catalyzed hydrolysis of RNA. Reaction products were resolved by denaturing PAGE and visualized by phosphorimaging. Reactions contained 0.1 μM ExoN (1:1 nsp14:nsp10) and 0.1 μM of the indicated RNA, were incubated at 30°C for the indicated time, then quenched by addition of EDTA. In both cases, primed-template-like dsRNA substrates were cleaved more efficiently than ssRNA substrates. **(E)** Exoribonuclease activity is dependent on the nsp14 active site. Residues of nsp14 required for catalysis were changed as follows: D90A, E92A; this derivative is referred to as MUT. Reactions contained 0.1 μM ExoN (1:1 nsp14(WT or MUT):nsp10) and 1 μM RNA and were run as described above. MUT did not exhibit any detectable exoribonuclease activity.

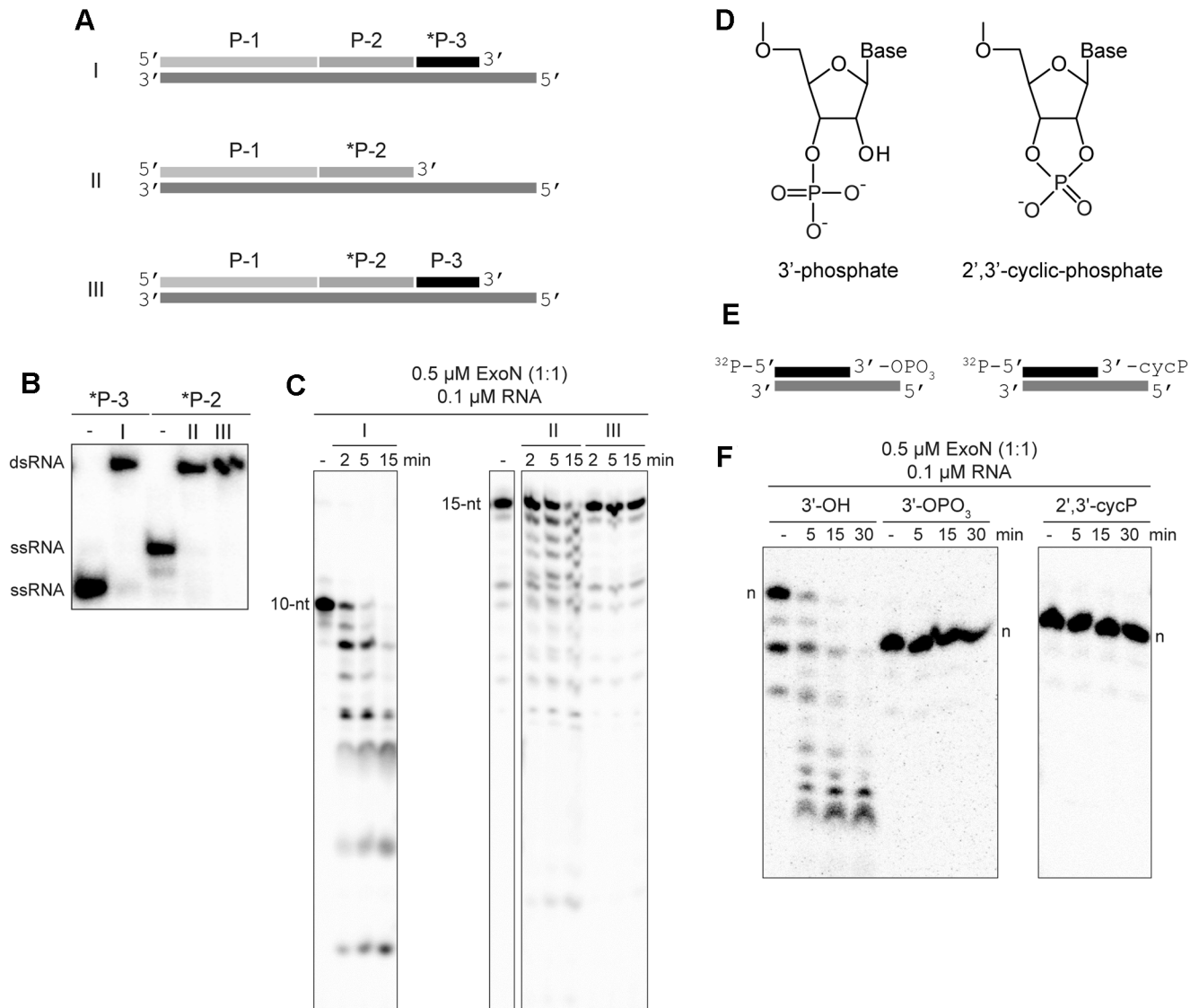


Figure 2. Products of nsp15-catalyzed endonucleolytic cleavage are not substrates for ExoN. (A) Schematic of dsRNA substrates used. Both substrates I and II contain 32 P-labeled RNAs (*P-3 and *P-2) annealed to the template such that both RNAs have an exposed 3'-end. Substrate III contains a 32 P-labeled RNA (*P-2) annealed to the template at a position where an additional RNA (P-3) is annealed downstream and blocks access to the 3'-end of the labeled RNA. Substrate III permits assessment of ExoN cleavage at a nick, as produced by nsp15 endonucleolytic cleavage. (B) RNA substrates evaluated by Native PAGE. Shown are the RNAs (*P-2 and *P-3) alone and when annealed to form the dsRNA substrates I, II or III. (C) ExoN does not initiate hydrolysis at a nick. The 10-nt and 15-nt RNAs are indicated. Reactions contained 0.5 μ M ExoN (1:1) and 0.1 μ M RNA and were quenched at the indicated times. The 32 P-labeled RNAs (*P-3 and *P-2) in substrates I and II that have an exposed 3'-end were efficiently cleaved by ExoN. The 32 P-labeled RNA (*P-2) in substrate III was not cleaved. (D–F) ExoN does not hydrolyze termini containing a 3'-phosphate or 2',3'-cyclic phosphate. The structures of these modifications are shown in panel D. Schematic of dsRNA substrates containing 3'-phosphate and 2',3'-cyclic phosphate modifications used are shown in panel E. *P10A:T21 dsRNA were used as substrates. Reactions contained 0.5 μ M ExoN (1:1) and 0.1 μ M RNA, were incubated for the indicated time, then quenched. Products are shown in panel F. The unmodified RNA is completely degraded; however, the 3'-phosphate and 2',3'-cyclic modifications block excision by ExoN. Note, the 3'-phosphate and 2',3'-cyclic modifications alter the apparent mobility of the RNA as it is more negatively charged and runs faster on the gel; the mobilities of each full-length RNA are indicated by n.

these conditions and upon dissociation, the product of one enzyme in the reaction immediately serves as a substrate for the same or a second enzyme in a reiterative manner until the product dissociates into two single-stranded molecules. This latter condition is the most prevalent in the literature (23,45–48,51,59).

Under both conditions, the earliest products were primarily $n - 1$ and $n - 2$ in length (Figure 3B). Quantitative analysis of both conditions showed consumption of >80%

of the substrate during each reaction (left panels of Figure 3C and D). More than 50% of the substrate was consumed by the first time point when enzyme was present in excess (left panel of Figure 3D), thus precluding an unambiguous assessment of precursor-product relationships (right panel of Figure 3D). However, under the condition of substrate-excess, the kinetics suggested production of $n - 1$ and $n - 2$ at the same rate, with $n - 2$ accumulating over the time course (right panel of Figure 3C). Therefore, products shorter than

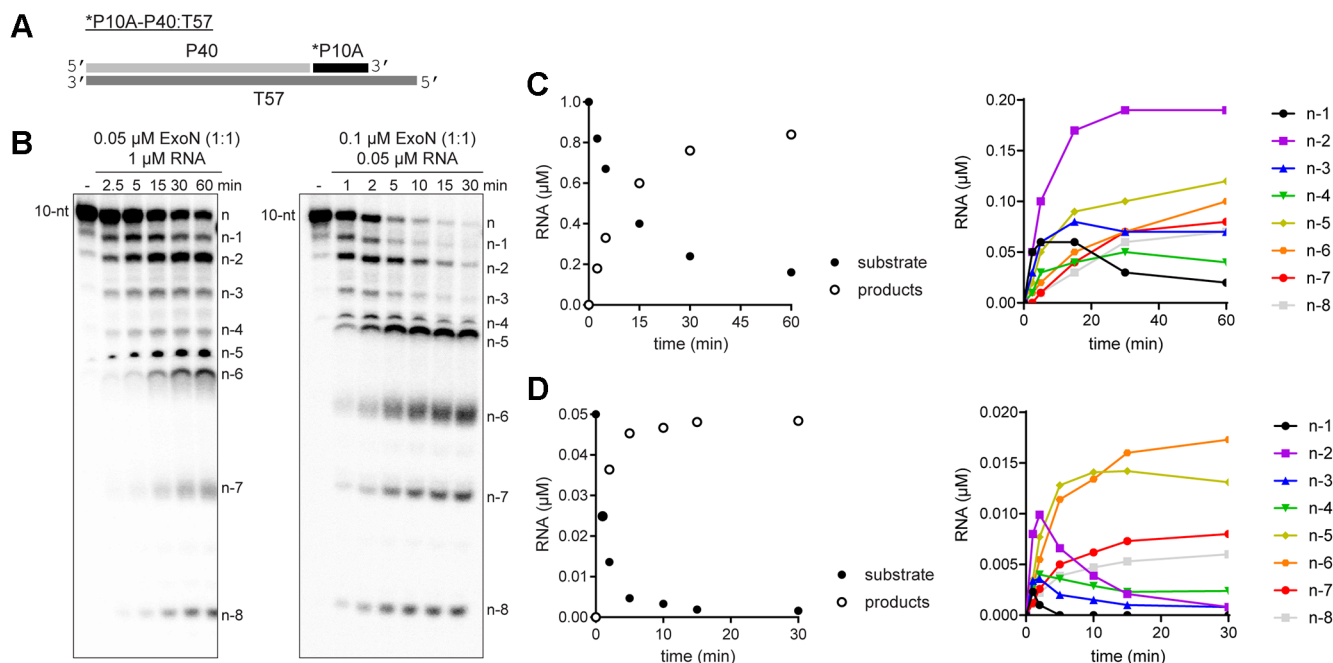


Figure 3. Monitoring ExoN-catalyzed RNA hydrolysis at single-nucleotide resolution. (A) Schematic of dsRNA substrate used to monitor ExoN activity at single-nucleotide resolution. (B) Reactions contained 0.05 μM ExoN (1:1) and 1 μM RNA or 0.1 μM ExoN (1:1) and 0.05 μM RNA, were incubated for the indicated time, then quenched. Products were resolved using denaturing 20 or 23% polyacrylamide gels. The RNA cleavage products, *n* – 1 to *n* – 8, are indicated. (C, D) Quantitative analysis of kinetics of substrate RNA utilization and/or product RNA formation. Formation of individual products is shown on the right. Panel C reports on the experiment in which RNA was in excess of enzyme; panel D reports on the experiment in which enzyme was in excess of RNA.

n – 3 likely arise from utilization of the *n* – 1 and/or *n* – 2 products.

Existence of an equilibrium between nsp14 and nsp10 proteins

Structural studies of ExoN demonstrate unambiguously that the nsp14:nsp10 stoichiometry is 1:1 (19,22,29,46,47,49). However, there has been an assumption that the affinity of nsp14 for nsp10 is sufficiently high that complex never dissociates once formed. This interpretation is based on the fact that most published studies emphasize the nsp14:nsp10 stoichiometry rather than the concentration of each component (20,23,29,45–52,59). Because most published studies have used conditions of enzyme excess, any dependence of the reaction rate on nsp10 concentration would be masked without evaluating product formation on the msec-sec timescale (Supplementary Figure S5).

We evaluated the reaction kinetics, processivity, and substrate specificity of nsp14 as a function of nsp10 concentration (Figure 4). The reaction rate increased as a function of nsp10 concentration, with apparent saturation at a concentration of 2 μM , and nsp14:nsp10 ratio of 1:20 (Figure 4). The primed-template-like dsRNA substrate was still favored over the ssRNA substrate (compare the 10-nt band in Figure 4A to that in Figure 4B). Finally, the distributive nature of the enzyme remained the same, with *n* – 1 and *n* – 2 products accumulating at early times for both substrates (Figure 4).

An understanding of the kinetics and thermodynamics of the binding reaction between nsp14 and nsp10 leading

formation of the nsp14-nsp10 complex is warranted. Such studies will require approaches to assess complex formation directly, rather than indirectly by monitoring exoribonuclease activity.

The presence of the phosphorothioate Rp diastereomer at the scissile phosphodiester bond blocks hydrolysis by ExoN

Polymerases prefer use of the Sp diastereomer when a phosphorothioate is placed at the alpha position of a nucleoside triphosphate (60–64). Incorporation yields the Rp diastereomer at the resulting phosphorothioate bond (Figure 5A), and the presence of this diastereomer can block hydrolysis by some exonucleases (65–69). We substituted the ultimate ($\text{P}_{\text{S}1}10\text{A}$ in Figure 5B), penultimate ($\text{P}_{\text{S}2}10\text{A}$ in Figure 5B), and antepenultimate ($\text{P}_{\text{S}3}10\text{A}$ in Figure 5B) phosphodiester bond with a phosphorothioate. Both diastereomers were present at a 50:50 ratio.

We assembled each phosphorothioate-substituted RNA into the **P10A-P40:T57* substrate to determine the impact of the phosphorothioate substitution on hydrolysis by ExoN compared to the control RNA (Figure 5C). Under the conditions of enzyme excess, essentially all of the control substrate was consumed (**P10A* in Figure 5C). However, only 50% or so of the substrate was consumed when a phosphorothioate was present in the RNA (**P_S10A*, **P_S210A* and **P_S310A* in Figure 5B), suggesting that one diastereomer inhibits hydrolysis. Importantly, the size of the terminal product in the reaction was consistent with the position of the phosphorothioate: *n*, for **P_S10A*; *n* – 1 for **P_S210A*; and *n* – 2, for **P_S310A* (Figure 5B).

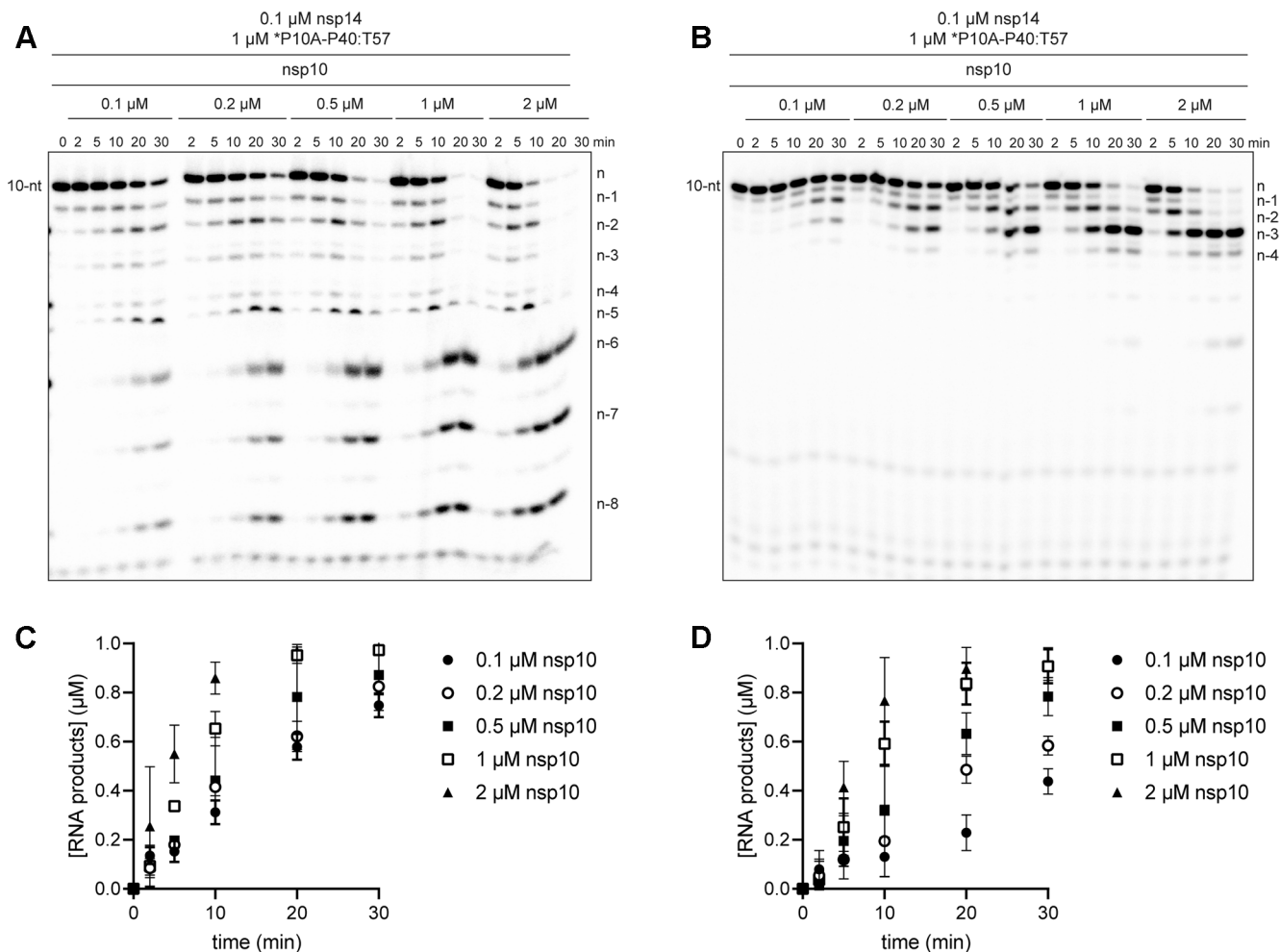


Figure 4. nsp10 stimulates hydrolysis without affecting processivity or substrate specificity. (A) Effect of nsp10 concentration on the kinetics of nsp14-catalyzed dsRNA hydrolysis. Reactions contained fixed nsp14 (0.1 μM) and varied nsp10 (0–2 μM) concentrations. In all cases, nsp14 was pre-mixed with nsp10 on ice 5 min prior to adding to the reaction. Hydrolysis of dsRNA substrate (*P10A-P40:T57) was monitored over a 30 min time course as indicated. Product analysis is shown. (B) Effect of nsp10 concentration on the kinetics of nsp14-catalyzed hydrolysis of ssRNA. Reactions were as in panel A except a ssRNA substrate (*P10A) was used. Product analysis is shown. Products shorter than $n-4$ are not observed when the 10-nt ssRNA is used. (C, D) Analysis of the kinetics of product formation for both dsRNA (panel C) and ssRNA (panel D) at varying concentrations of nsp10, error bars indicate mean \pm SD ($n = 3$).

To determine if the phosphorothioate Rp diastereomer was the inhibitory species, we used polymerase incorporation of nucleoside-5'-O-(1-thiotriphosphates) to produce RNA with only the Rp diastereomer at the scissile bond (Figure 6A). For these experiments, we used the cryptic RdRp activity present in the mitochondrial DNA-dependent RNA polymerase (DdRp), POLRMT. For these experiments, we used the DdRp to produce an RNA product with the terminal scissile bond containing the phosphorothioate Rp diastereomer (Figure 6B). After $\sim 50\%$ of the primer was extended to product, we added ExoN and monitored the fate of the product RNA by denaturing PAGE and phosphorimaging (Figure 6B). Extension of only half of the product provides an internal control for the presence of ExoN, because the primer lacks a phosphorothioate bond. We used two different substrates. For both, the presence of the phosphorothioate Rp isomer completely blocked hydrolysis by ExoN (right panels in Figure 6C, D), relative to both the unextended primer (internal control) and the com-

parable RNA product lacking the phosphorothioate (left panels in Figure 6C, D). Quantitation is provided in Figure 6E and F. Addition of the phosphorothioate to the α -position of an antiviral nucleotide analog should eliminate the possibility for excision by ExoN.

Composition of the terminal basepair modulates the kinetics of hydrolysis by ExoN

For proofreading exonucleases of DNA polymerases, it is not that the exonuclease exhibits a preference for a mismatched end relative to a properly paired end (27,28). Rather, a mismatched end is unstable in the polymerase active site, providing the opportunity for the mismatched end to partition to the exonuclease active site (27,28). To assess the impact of the terminal basepair on the kinetics of excision by ExoN, we designed two additional RNA substrates for ExoN. *P10G creates a terminal G:U mismatch; *P10R creates a terminal remdesivir:U pair, which is equivalent to an

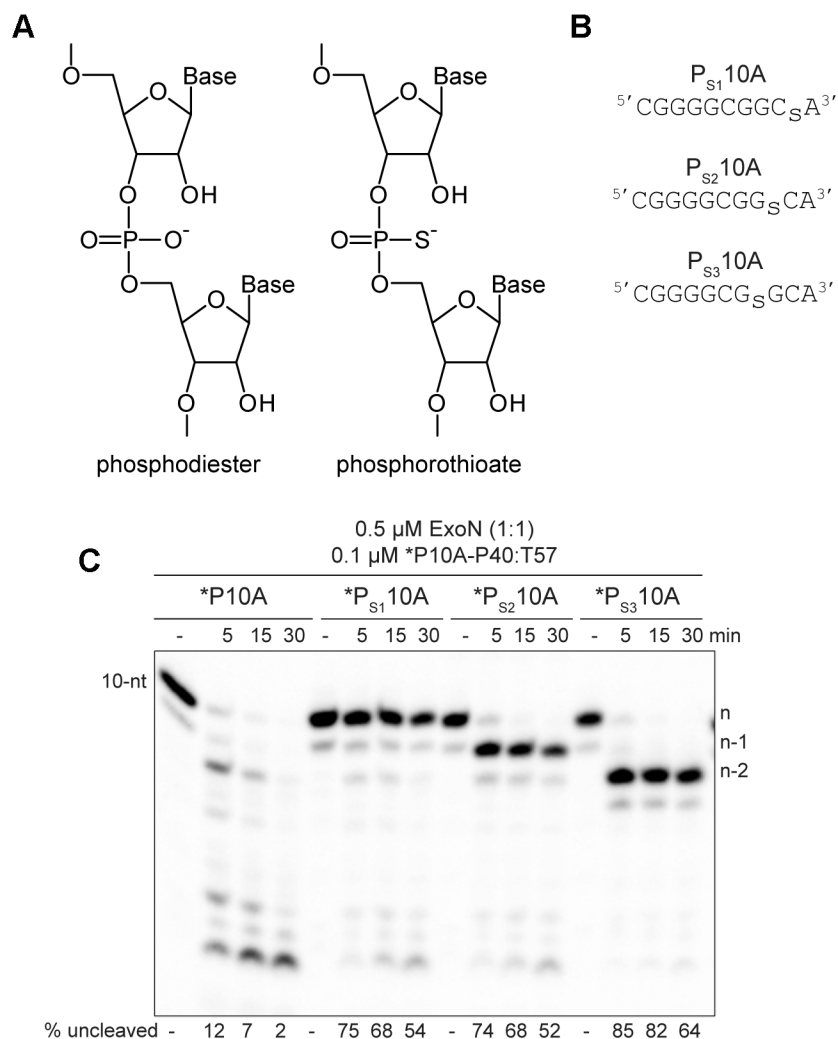


Figure 5. RNA hydrolysis by ExoN is inhibited by a phosphorothioate bond. (A) Structure of phosphorothioate. One of the non-bridging oxygens of the phosphodiester bond is replaced with a sulfur atom, creating Rp and Sp diastereomers. A 50:50 ratio of diastereomers were used in these experiments. (B) Schematic of phosphorothioate-substituted ssRNAs used. (C) Effect of phosphorothioate ExoN-catalyzed hydrolysis of dsRNA. Reactions contained 0.5 μM ExoN (1:1) and 0.1 μM of the indicated RNA, were incubated at the indicated times, then quenched. Product analysis is shown. Fraction of RNA remaining is indicated (% uncleaved). The unmodified RNA is completely degraded; however, ExoN is unable to hydrolyze beyond the phosphorothioate substitution, for half the RNA molecules at least. This observation is consistent with only a single diastereomer being inhibitory.

A:U pair at the level of the hydrogen bonding potential between the bases (Figure 7A). We evaluated the kinetics of hydrolysis of the various RNA substrates by ExoN under conditions in which the RNA substrate was present in excess of enzyme. Reaction products were resolved by denaturing PAGE and visualized by phosphorimaging (Figure 7B). The observed rates of consumption of the RNA substrate and formation of all products were the same (Figure 7C). Any difference that would be observed, however, would be related to hydrolysis of 3'-terminal nucleotide to produce the $n-1$ product and any subsequent excision reactions occurring prior to dissociation of ExoN. We performed a comprehensive analysis of the kinetics of formation of each product for each substrate (Figure 7D). We observed a clear difference in the magnitude of the $n-1$ product that accumulated for each substrate (Figure 7D). Such an observation would suggest that the rate of the step governing hydrolysis and/or rate of dissociation of the excised nucleotide

product differ between the substrates. Interestingly, excision of remdesivir was substantially slower than the natural nucleotides as the $n-1$ product accumulated to the greatest extent for the P10R substrate (Figure 7D).

Some modifications to the 3'-terminal ribose antagonize ExoN-catalyzed excision

To further our understanding of the structure–activity relationships of the 3'-terminal nucleotide, we prepared substrates in which the 2'- or 3'-hydroxyl was changed (Figure 8A). Our standard substrate RNA is P10A. Unfortunately, not all modifications were available as an adenosine phosphoramidite. Therefore, some modifications were analyzed in the context of P10U (Figure 8A and Supplementary Figure S7). We monitored the kinetics of ExoN cleavage of modified RNAs. Products were resolved by denaturing PAGE and visualized by phosphorimaging (Supplementary

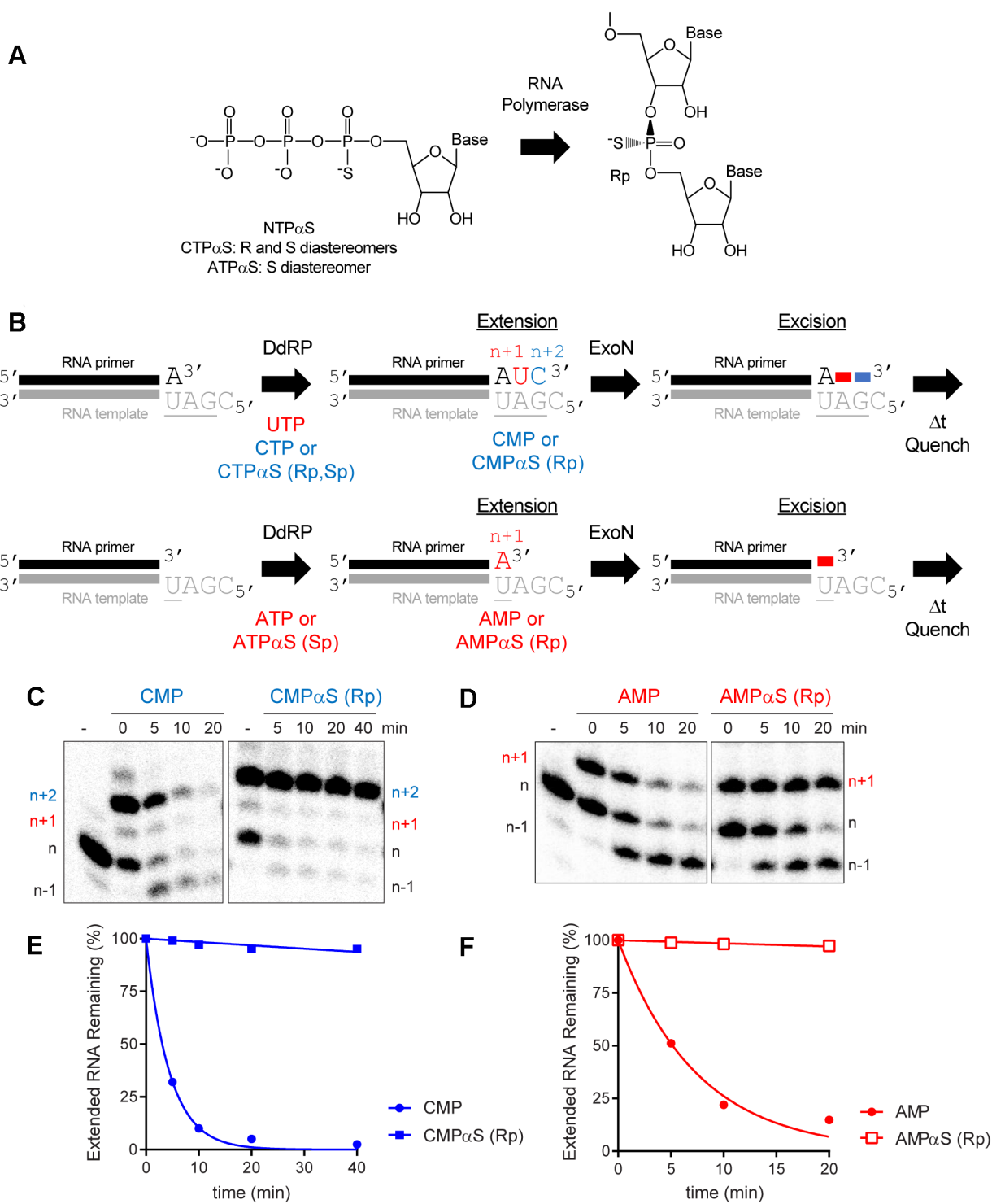


Figure 6. ExoN is inhibited by the Rp diastereomer at the phosphorothioate-substituted scissile bond. (A) The Sp diastereomer of a nucleoside-5'-O-thiotriphosphate is preferentially incorporated by RNA polymerases. Nucleophilic attack at the alpha phosphorous atom leads to inversion of configuration and creation of Rp diastereomer at the phosphorothioate bond. (B) Schematic of assay. Primer extension is initiated by adding an RNA polymerase in the presence of either UTP and CTPαS or ATPαS. CTPαS is a mixture of both the Rp and Sp diastereomers. ATPαS was obtained as the pure Sp diastereomer. Incorporation yields extended products. The last nucleotide to be incorporated is either CMP(αS) or AMP(αS). Once 50–75% of the primers were extended to the end, ExoN was added to the reaction. The reaction was monitored over time for hydrolysis. Unextended primer in reactions served as a useful control to demonstrate the presence of active ExoN in the reaction. P9:T21 was used as dsRNA substrate for these reactions. (C, D) Analysis of reaction products by denaturing PAGE. Incorporation of CMP and AMP results in removal of the incorporated nucleotide by ExoN. Incorporation of CMPαS and AMPαS (Rp diastereomers) results in a terminated primer that cannot be cleaved by ExoN. (E, F) Kinetics of excision of CMP, CMPαS (Rp diastereomer) and AMP and AMPαS (Rp diastereomer) by ExoN. Data were fit to a single exponential. Rates are provided in Table 2.

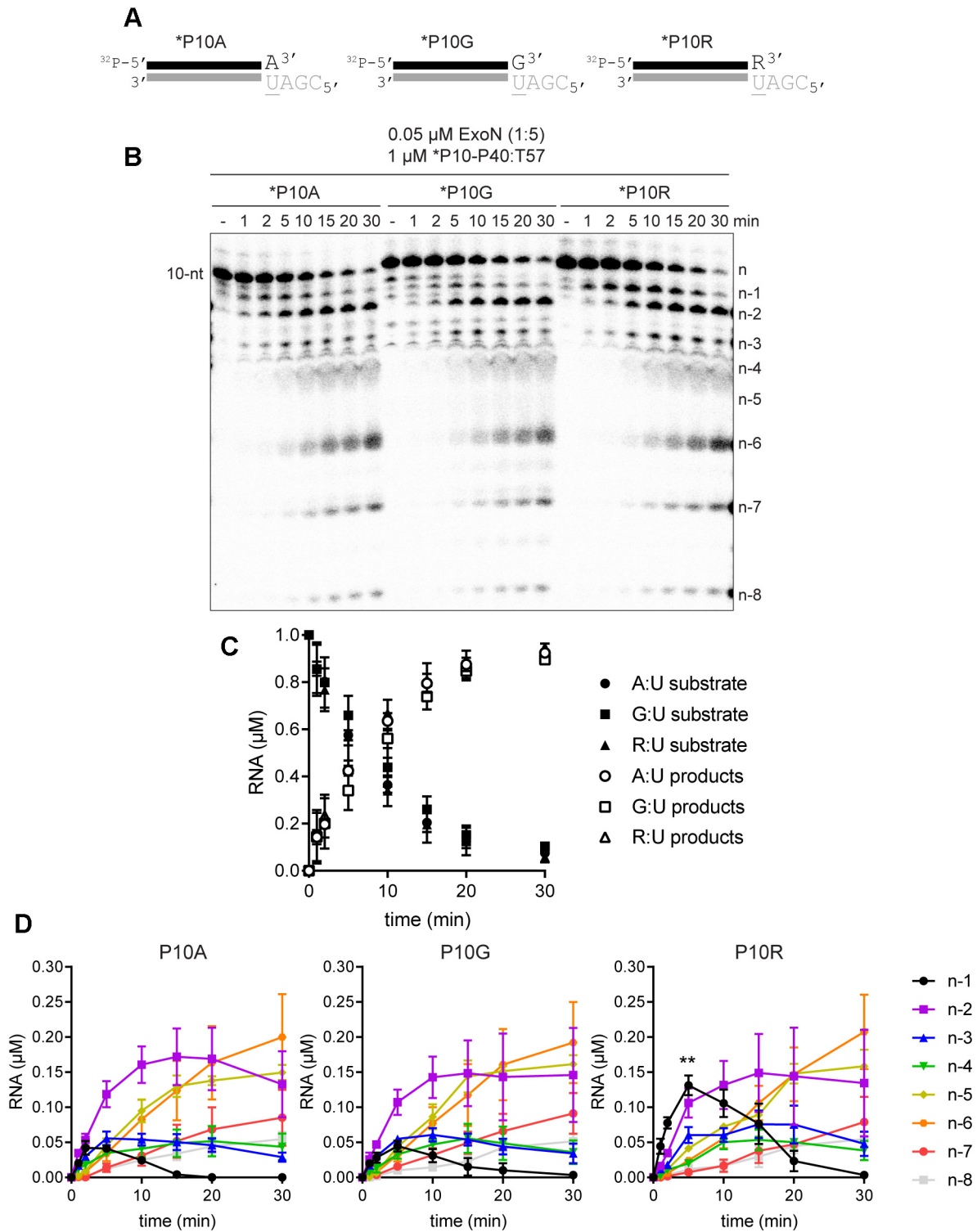


Figure 7. Terminal basepair-dependent differences in the kinetics of hydrolysis by ExoN. **(A)** Schematic of dsRNA substrates used to evaluate the influence of the terminal basepair on kinetics of excision by ExoN. The termini have a A:U, G:U or R:U basepair, where R represents remdesivir. **(B)** Effect of terminal basepair on hydrolysis by ExoN. Reactions contained 0.05 μM ExoN (1:5) and 1 μM of the indicated dsRNA substrate, were incubated at 30°C for the indicated times, then quenched. Product analysis is shown. **(C)** Analysis of the kinetics of substrate utilization and product formation. Here, depletion of substrate and total product formation are monitored, showing no difference in the kinetics, error bars indicate mean ± SD ($n = 3$). **(D)** Analysis of the kinetics of formation of each individual product. Cleavage products ($n - 1$ to $n - 8$) were plotted as a function of time for each RNA substrate, error bars indicate mean ± SD ($n = 3$). The kinetics of formation and utilization of the $n - 1$ product vary between substrates, suggesting that some, but not all steps of the mechanism are agnostic to the sequence and/or complementarity of the terminal basepair. Statistical analysis (unpaired t -test) comparing the $n - 1$ products from P10A to P10R at 5 min is indicated, $P = 0.0038$.

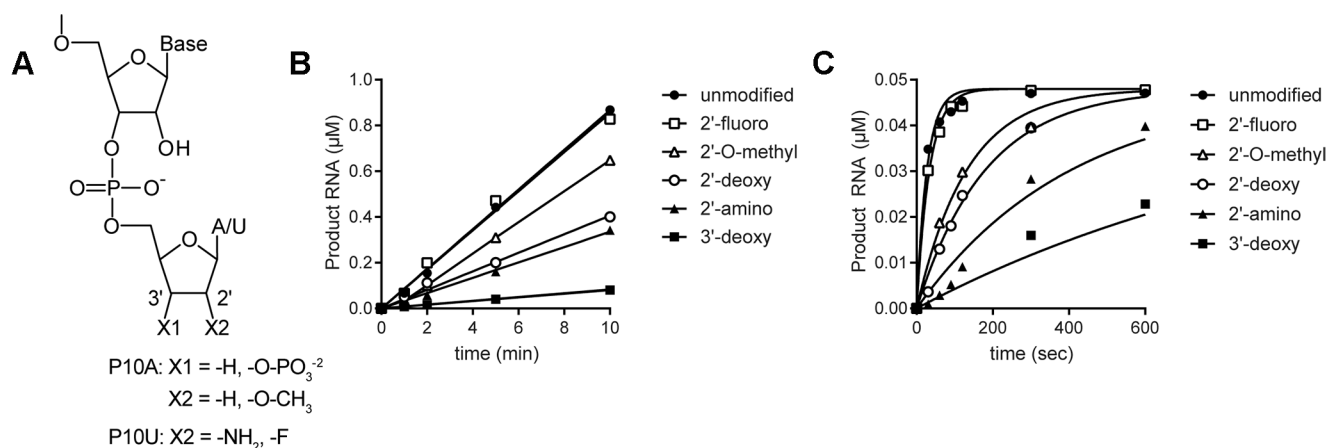


Figure 8. Structure-activity relationships for the ribose of the 3'-terminal nucleotide. (A) Modifications to the ribose of the terminal nucleotide studied here in the context of P10A: 3'-deoxy (-H), 3'-phosphate (-O-PO₃), 2'-deoxy (-H) and 2'-O-methyl (-O-CH₃); or P10U: 2'-amino (-NH₂) and 2'-fluoro (-F). (B, C) Kinetics of cleavage of modified dsRNAs. In panel B, substrate was present in excess of enzyme. In panel C, enzyme was present in excess of substrate. The data were fit to a line (panel B) or to a single exponential (panel C). Reactions contained 0.1 μM ExoN (1:20) and either 1 or 0.05 μM dsRNA substrate and were quenched at the indicated times. The order in which the modified RNAs were cleaved is as follows: unmodified = 2'-F > 2'-O-Me > 2'-deoxy > 2'-amino > 3'-deoxy > 3'-phosphate. The kinetics of cleavage using the 3'-phosphate RNA is not shown, because cleavage was not detected. The observed rate of cleavage of modified RNAs and fold difference from the unmodified RNA are shown in Table 3.

Figures S6 and S7). We plotted the concentration of total RNA products as a function of time for the two conditions used: excess substrate (Figure 8B) or excess enzyme (Figure 8C). The corresponding observed rates are presented in Table 3. The presence of a 3'-phosphate blocked hydrolysis (Figure 2E). The presence of a 2'-fluoro had no effect. All other substitutions reduced the efficiency of cleavage, with the most effective being 3'-H, 2'-NH₂ and 2'-H (Figure 8B, C and Table 3). Together, these data suggest that it may be possible to identify substituents on the 3'-terminal ribose that would not impact utilization by the RTC but would inhibit excision by ExoN.

ExoN efficiently excises the chain-terminating antiviral ribonucleotide produced by viperin, ddhC

The literature is rife with examples of the inability of obligate and non-obligate chain terminators to interfere with SARS-CoV-2 multiplication in cell culture (8,13,70–72), even the antiviral ribonucleotide, ddhCTP, produced by the antiviral protein, viperin (73). However, a 3'-dAMP terminated RNA is not a good substrate for ExoN (Figure 8), and others have made similar observations for other 3'-dNMPs (29,74).

Given this apparent contradiction between the cell-based and biochemical experiments, we asked how ExoN would deal with the presence of the viperin product, ddhCTP, and the non-obligate chain terminator, 2'-C-Me-CTP (structures shown in Figure 9A), when present at the 3'-end of RNA. Because synthetic RNA containing these analogs was not available, we made these RNAs biosynthetically by taking advantage of the cryptic RdRp activity present in the mitochondrial DNA-dependent RNA polymerase (DdRp, aka POLRMT, Figure 9B). CTP and 3'-dCTP were used as controls to permit comparison to data obtained using synthetic RNA. After production of the CMP analog-terminated RNA, ExoN was added, and the reaction proceeded for the indicated time before quenching (Figure 9B).

Reaction products were resolved by denaturing PAGE and visualized by phosphorimaging (Figure 9C) and quantified (Figure 9D). The presence of the 2'-C-Me substituent had little to no effect on ExoN activity (2'-C-Me-CTP in Figure 9C, D). In spite of the absence of a 3'-OH on ddhCTP, ddhCMP was excised from RNA almost as efficiently as CTP (ddhCTP in Figure 9C, D). Together, the results presented in Figures 8 and 9 reveal complexity to the interaction between the 3'-terminal nucleotide and the ExoN active site that is not explained by simple docking experiments. A more in-depth characterization of the substrate specificity of ExoN is warranted.

SARS-CoV-2 RTC prevents access of ExoN to the 3'-end nascent RNA

We have been able to assemble elongation complexes comprised of SARS-CoV-2 RTC and our split-primer:template substrate used for ExoN (Figure 10A). Studies using similarly designed substrates with shorter duplexes fail to assemble a stable complex (data not shown). The ability to split the primer provides the advantage of being able to monitor extension at single-nucleotide resolution.

Assembly and incubation of a complex with a correctly basepaired 3'-end for 60 min (Figure 10A; 0 min Figure 10B) remains stable to challenge with ExoN for at least an additional 60 min (Figure 10A; CTP in Figure 10B, C). Judicious selection of substrate nucleotides and/or analogs permitted us to exploit this system to create elongation complexes with different 3'-ends. This experimental design permitted us to determine if the nature of the 3'-terminal base-pair under investigation influenced the stability of assembled SARS-CoV-2 RTC and/or accessibility of ExoN. Neither a mismatched end (UTP in Figure 10B, C) nor an end containing a chain terminating analog (2'-C-Me-CTP, 3'-dCTP, ddhCTP in Figure 10B, C) changed the sensitivity to ExoN, so by inference did not change stability with the SARS-CoV-2 RTC either. Extension of shorter duplexes by

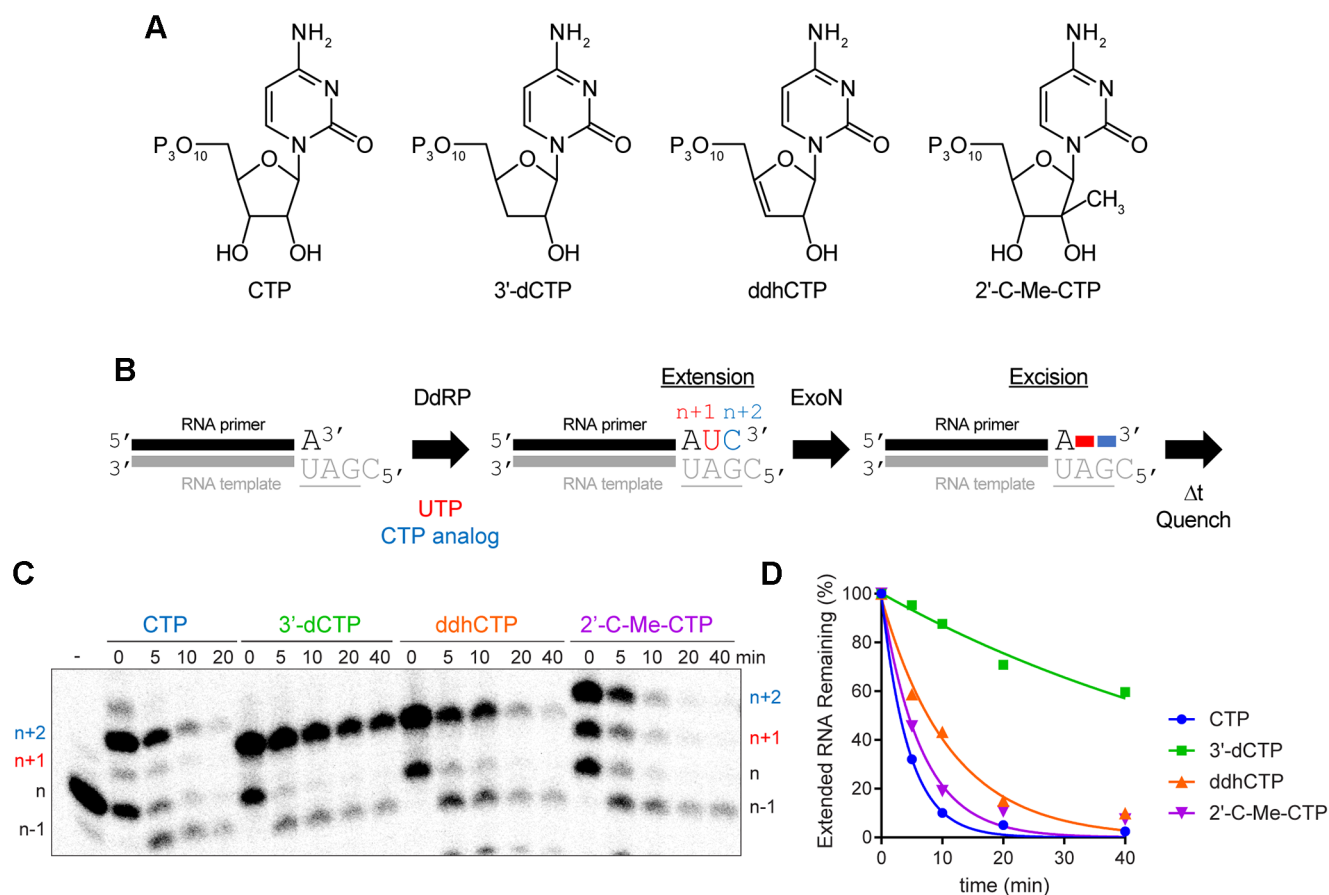


Figure 9. Viperin product (ddhCMP)-terminated RNA is a good substrate for ExoN but 3'-dCMP-terminated RNA is not. (A) Structure of nucleotide analogs used. (B) Schematic of the assay. Primer extension is initiated by adding an RNA polymerase in the presence of UTP and a CTP analog. Incorporation produces $n + 1$ and $n + 2$ products. The last nucleotide to be incorporated is the CTP analog. Once 50–75% of the primers were extended to $n + 2$, ExoN was added to the reaction. The reaction was monitored over time for hydrolysis. Unextended primer in reactions served as a useful control to demonstrate the presence of active ExoN in the reaction. (C) Analysis of reaction products by denaturing PAGE. Only the 3'-dCMP-terminated RNA exhibited a delay in excision. (D) Kinetics of excision of CMP analogs. The quantitation revealed that only 3'-dCMP-terminated RNA exhibited a significant delay in the rate of excision relative to the CMP control, even though ddhCMP also lacks a 3'-OH. Data were fit to a single exponential. Rates are provided in Table 4.

the SARS-CoV-2 RTC that failed to form a stable complex were excised by ExoN (data not shown).

Together, these data show that the SARS-CoV-2 RTC creates a physical block to cleavage and that the composition of the terminal basepair does not trigger a response by the SARS-CoV-2 RTC that facilitates access by ExoN.

DISCUSSION

Coronaviruses encode a 3'-5' exoribonuclease (11,13–15). A complex of the nsp10 and nsp14 proteins at a stoichiometry of 1:1 forms the active enzyme (19,22,29,47,49). The nsp14 subunit harbors the catalytic residues; however, this subunit fails to achieve a stable conformation competent for RNA binding and catalysis in the absence of nsp10 (22). Therefore, we have referred to the active complex as ExoN instead of the nsp14 subunit alone. The biological function of ExoN is proofreading, correction of errors introduced by the viral polymerase (11–14). ExoN also contributes to excision of antiviral nucleotides, as the sensitivity of the virus to certain antiviral nucleotides increases in the absence of exoribonuclease activity (25). This latter activity complicates the

use of some conventional antiviral nucleotides for the treatment of coronavirus infection. The goal of this study was to discover strategies to interfere with excision of nucleotides by ExoN and thereby inform the development of nucleotide analogs with greater efficacy against coronaviruses.

The catalytic cycle of ExoN includes binding of the RNA substrate to the enzyme followed by hydrolysis (Figure 11A). There is now a consensus that the RNA substrate preferred by ExoN is one that would also be preferred by the replication-transcription complex (RTC), a primed-template (Figure 1) (20,29,49). The number of nucleotides hydrolyzed per binding event was not known before this study. This gap existed for two reasons. First, the design of the RNA substrates used in these studies made difficult – if not impossible – the monitoring of product formation with single-nucleotide resolution (23,29,45,47–51,59). Second, ExoN was present in substantial excess over the RNA substrate, thus requiring monitoring of hydrolysis on the millisecond timescale to observe a product one nucleotide shorter than the substrate, which was never done (23,29,45,47–51,59). The inability of ExoN to hydrolyze nucleotides at a nick permitted us to distribute

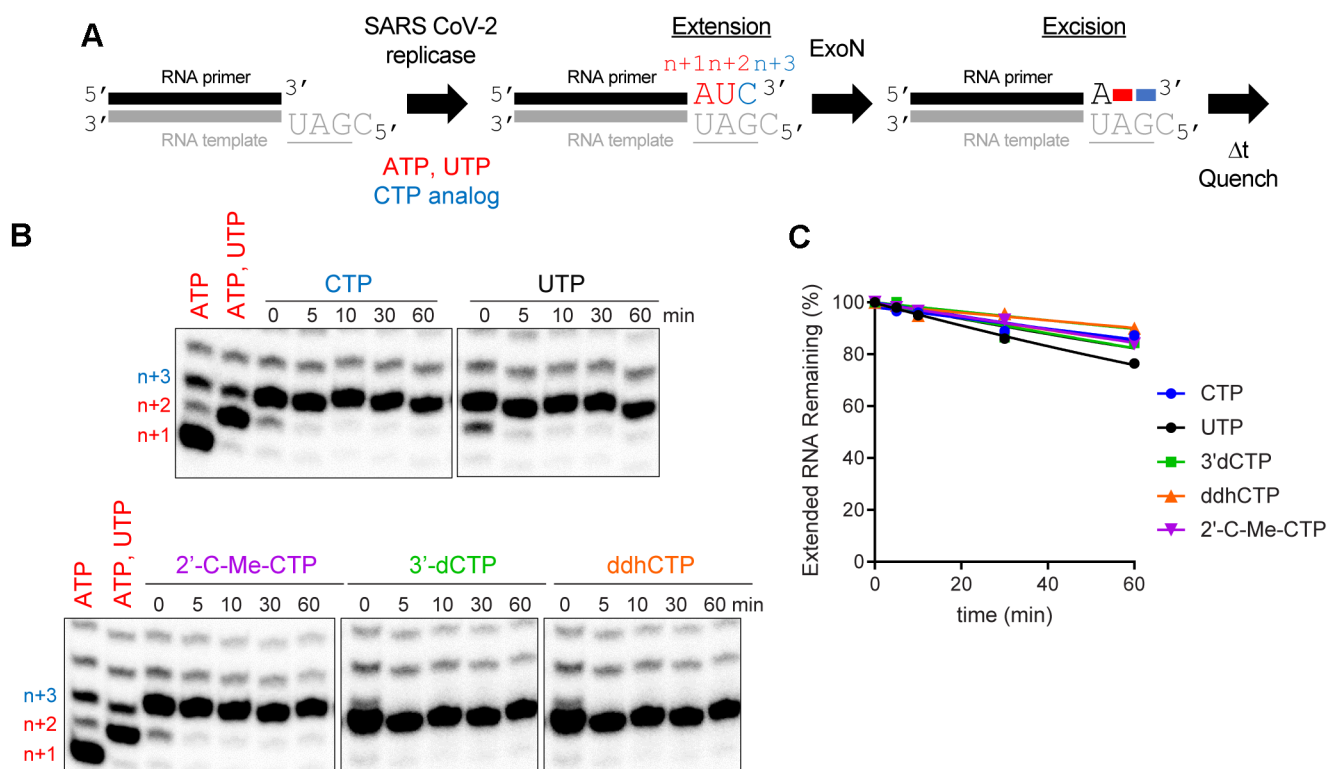


Figure 10. Stalled SARS-CoV-2 replication-transcription complexes block access of ExoN to the RNA terminus. (A) Schematic of assay. Primer extension was initiated by adding the SARS-CoV-2 replicase-transcription complex (nsp7/nsp8/nsp12) in the presence of ATP and UTP or ATP, UTP and CTP or a CTP analog. After an incubation time of 60 min, we added ExoN, incubated for the indicated time, and then quenched the reaction. (B) Effect of ExoN on stalled elongation complexes. Incorporation yields $n + 1$, $n + 2$ and $n + 3$ products. When ATP, UTP and CTP are present, the $n + 3$ product has a properly basepaired 3'-end. The $n + 4$ product is formed by misincorporation and has a mispaired 3'-end. When only ATP and UTP (low concentration, 1 μ M UTP) are present, both the $n + 3$ and $n + 4$ products are formed by misincorporation and both products have mispaired 3'-ends. At high concentrations of UTP (100 μ M), most of the products observed are at $n + 3$ from misincorporation of U opposite G. ExoN was added to elongation complexes for the indicated time (5–60 min). Reaction products were visualized by phosphorimaging after denaturing PAGE. No change in the level of properly paired ($n + 3$) or mispaired ($n + 4$) elongation products was apparent for any of the reactions performed, consistent with the replicase remaining bound to both products and obstructing access by ExoN. (C) Quantitation of reaction products shown in panel B. The data were fit to a line. Rates are provided in Table 4.

the overall requirement for dsRNA over multiple segments, which enabled product analysis at single-nucleotide resolution (Figure 2A, B). While performing the experiments under conditions of enzyme excess may have some utility when evaluating inhibitors (45), only conditions of substrate excess can provide information on the number of nucleotides hydrolyzed per binding event when quenching reactions manually (Figure 11A). ExoN excised only one or two nucleotides per binding event (Figure 3). The distributive nature of ExoN makes sense, because a processive enzyme would then require the RTC to resynthesize RNA that lacked any damage.

For a nucleotide analog to resist excision, the rate constant for excision (Step 2 in Figure 11A) needs to be slower than the rate constant for dissociation of ExoN from the 3'-terminus (Step 4 in Figure 11A). RNA product dissociation is likely the rate-limiting step for this reaction under steady-state conditions (Figure 11A). Both the 2'- and 3'-hydroxyls contribute to the efficiency of excision, as modifications to either appear to reduce the rate of catalysis to levels at or below the rate constant for dissociation (Figure 8, Table 3). Our assumption is that when the rate of product formation is the same under conditions of excess sub-

strate and excess enzyme, then the rate measured is catalysis (Table 3). The 3'-position was most important. Loss of the 3'-OH made catalysis rate limiting (Figure 8, Table 3); addition of a 3'-phosphate eliminated turnover altogether (Table 3). The SARS-CoV-2 RTC has no problem utilizing nucleotide substrates lacking a 3'-OH (8,75). The issue with most analogs is the inability to compete with natural nucleotide pools (8,75). Remdesivir does not suffer this problem (8,75). Perhaps elimination of its 3'-OH will bolster its anti-coronavirus activity.

The ability of 3'-dCMP to antagonize ExoN-mediated hydrolysis so effectively was not expected, although several other groups have now reported this observation (29,74). Our previous studies showed that the antiviral nucleoside, ddhC, does not interfere with SARS-CoV-2 multiplication in cells, even though the RTC utilized the nucleotide, ddhCTP, quite readily (8). ddhC lacks a 3'-OH, but ExoN readily excised ddhCMP from the 3'-end of RNA (Figure 9). We conclude that the requirement for a 3'-OH is not absolute but is context dependent. The context is likely associated with the ribose conformation. The ribose component of normal nucleotides are not planar and possess 'sugar pucker' with the 3'-endo conformation being favored

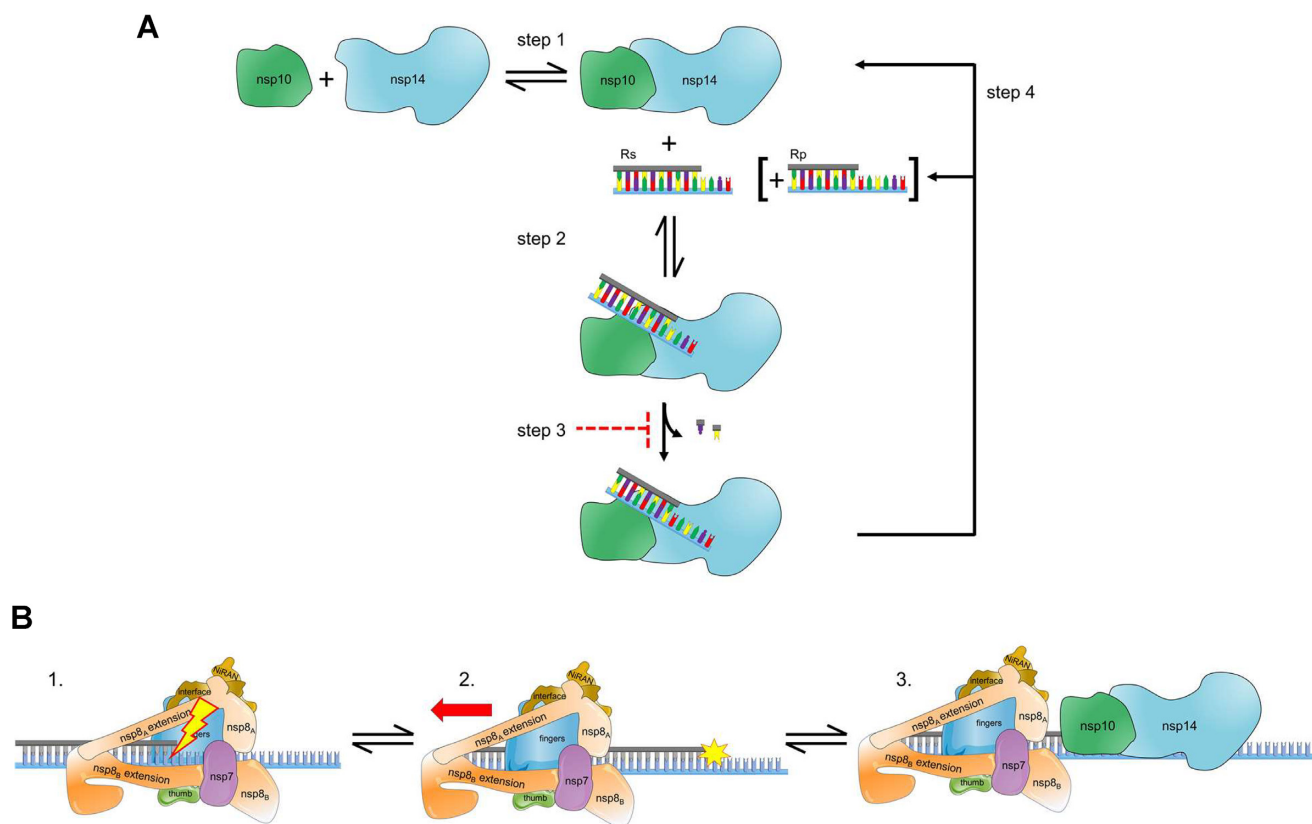


Figure 11. Hypothetical model for ExoN-catalyzed excision. **(A)** ExoN uncoupled to RNA synthesis. The active exoribonuclease, referred to here as ExoN, is a complex of two subunits: nsp10 (regulatory) and nsp14 (catalytic). The concentration of ExoN therefore depends on the value of the dissociation constant for this binding reaction, as well as the concentrations of nsp10 and nsp14. Most studies to date have ignored this equilibrium. Under steady-state conditions in which the RNA substrate (R_S) is present in excess of ExoN, R_S binds, catalysis occurs, releasing one or two nucleotides, followed by dissociation of the RNA product (R_P). Under these conditions, the rate-limiting step is likely release of R_P . Modifications to the terminal nucleotide that reduce the rate of hydrolysis to values substantially lower than the rate of R_P dissociation should interfere with excision. R_P consumption depends on its ability to compete with R_S , for example when more than 20% of the substrate is consumed or when ExoN is present in excess of substrate and more than a single turnover occurs. The possibility also exists for product dissociation to be driven by dissociation of the nsp10 and nsp14 subunits. **(B)** ExoN coupled to RNA synthesis. Introduction of mismatched nucleotides or chain terminators by the polymerase complex does not lead to dissociation, blocking access of the 3'-end for repair by ExoN. The polymerase complex must therefore be actively dislodged from the terminus to access the 3'-end. Once ExoN binds, only one or two nucleotides will be removed.

Table 2. Kinetics of excision of the R_P diastereomer of $CMP\alpha S$ and $AMP\alpha S$ by ExoN. Data were fit to a single exponential to determine the observed rate of cleavage and half-life ($t_{1/2}$)

NMP	Rate (min^{-1})	$t_{1/2}$ (min)
CMP	0.22 ± 0.010	3
$CMP\alpha S$ (R_P, S_P)	0.0016 ± 0.0003	350
AMP	0.13 ± 0.02	5
$AMP\alpha S$ (R_P)	0.0015 ± 0.0003	460

for ribonucleosides. The ribose of ddhC on the other hand is forced into a planar conformation because of the double bond between carbons 3' and 4' (Figure 9). Structural studies will be required to provide an explanation for the context-dependent requirement for the 3'-OH.

In retrospect, it is not surprising that coronaviruses have evolved to evade the inhibitory activity of ddhCMP-terminated RNA. ddhCTP is induced by infection (73,76). Indeed, the presence of ddhCTP in cells presents a strong selective pressure for acquisition of an exoribonuclease, es-

Table 3. Effect of modifications to the 3'-end on the kinetics of excision by ExoN. Data were fit either to a line ($S > E$) or to a single exponential ($E > S$) to determine observed rate of cleavage

Modification	Rate (min^{-1}) $S > E$	Rate (min^{-1}) $E > S$
Unmodified	0.87 ± 01	2.2 ± 0.2
2'-fluoro	0.85 ± 0.3	1.7 ± 0.1
2'-O-methyl	0.64 ± 0.02	0.46 ± 0.04
2'-deoxy	0.41 ± 0.01	0.33 ± 0.02
2'-amino	0.33 ± 0.01	0.15 ± 0.02
3'-deoxy	0.035 ± 0.006	0.056 ± 0.008
3'-phosphate	No cleavage	No cleavage

pecially given the efficiency with which the RTC utilizes this antiviral nucleotide (8). The presence of an exoribonuclease may have also contributed to the polymerase retaining the ability to utilize ddhCTP. The evolutionarily related enteroviral polymerases have lost the ability to utilize ddhCTP (73). Some coronaviruses deal with loss of the exoribonuclease better than others(23), perhaps the efficiency with which

Table 4. Kinetics of excision of chain terminating cytidine analogs by ExoN. Data were fit to a single exponential to determine the observed rate of cleavage and half-life ($t_{1/2}$)

NTP	-SARS replicase ^a		+SARS replicase ^b	
	Rate (min ⁻¹)	$t_{1/2}$ (min)	Rate (min ⁻¹)	$t_{1/2}$ (min)
CTP	0.23 ± 0.01	3	0.0027 ± 0.0004	260
UTP	-	-	0.0046 ± 0.0002	150
3'-dCTP	0.014 ± 0.002	50	0.0032 ± 0.0004	220
ddhCTP	0.089 ± 0.01	8	0.0018 ± 0.0002	390
2'-C-Me-CTP	0.15 ± 0.02	5	0.0028 ± 0.0001	250

^aExperiment shown in Figure 9.^bExperiment shown in Figure 10.

the polymerases utilize ddhCTP has something to do with this. It would be interesting to determine the requirement of the exoribonuclease in cells lacking viperin, the enzyme that produces ddhCTP.

The most efficient strategy to interfere with nucleotide excision is to introduce a non-hydrolyzable scissile bond. Phosphorothioate (P-S) substitutions at the scissile phosphodiester bond inhibit excision by DNA exonucleases, usually requiring a specific stereoisomer (S_P or R_P) (65–69,77). Here we show that the presence of a phosphorothioate in the R_P configuration of the scissile phosphodiester bond inhibits hydrolysis by ExoN (Figures 5 and 6 and Table 2). The ability to block cleavage of the terminal nucleotide rules out the existence of ExoN-associated endonuclease activity as suggest recently (45). Introduction of the P-S-substituted bond at ultimate, penultimate, and/or antepenultimate position will create substrates to facilitate elucidation of the details of ExoN-catalyzed hydrolysis by limiting the extent to which a product can be consumed. Current strategies to synthesize P-S-substituted oligonucleotides yield diastereomeric mixtures at phosphorous, but stereoisomer-specific solutions are on the horizon (77).

Another interesting possibility for use of the P-S substitution is the creation of non-hydrolyzable antiviral ribonucleotide analogs. Most clinically used monophosphorylated ribonucleotide analogs are readily delivered to cells as prodrugs that require intracellular activation by histidine triad nucleotide-binding protein 1 (HINT-1) (78,79), the so-called ProTide strategy (7,80–82). Similar strategies can be used for P-S-substituted ribonucleoside monophosphates. It is clear that HINT-1 can activate such compounds in cells, although the kinetics of activation are slower than conventional ProTides (78). HINT-1 is known to convert P-S-substituted NMPs to the natural NMP (78,79). The efficiency of this side reaction is likely tunable as the nature of the base and sugar contribute to the rate of sulfur removal (78). Side reactions such as these could also be alleviated by using P-S-substituted di- or triphosphorylated prodrugs (83). The pursuit of non-hydrolyzable ribonucleotide analogs is a guaranteed solution to the problem of excision of antiviral ribonucleotides by ExoN.

The majority of our studies focused on excision independent of RNA synthesis. An early report on ExoN from SARS-CoV suggested that co-transcriptional repair could occur in vitro (46,48). These studies used primed-templates of insufficient length to form a stable elongation complex (46,48). So, just as we were able to see ‘co-transcriptional’

repair using POLRMT (e.g. Figure 9), this repair required dissociation of POLRMT to permit ExoN access to the 3'-end. Formation of a stable elongation complex using the SARS CoV-2 RTC requires a duplex of 50-bp or greater (30,57,58). The ability to split the primer into 40-nt and 10-nt fragments permitted us to monitor extension by the RTC and excision by ExoN at single-nucleotide resolution in the same reaction (Figure 10). Interestingly, assembling the SARS CoV-2 RTC on these primed templates and forcing the complex to make errors was insufficient to get the complex to dissociate and expose the 3'-end to ExoN (Figure 10). ExoN removed these same termini in the presence of POLRMT (Figure 9).

We conclude that the SARS-CoV-2 RTC alone does not sense the presence of a mismatch, a non-natural nucleotide, or the pause caused by a chain terminator (panel 1 in Figure 11B). A more active mechanism must exist to remove the RTC from the 3'-end, perhaps another viral factor pushes the RTC out of the way (panel 2 in Figure 11B). The nsp13-encoded 5'-3' RNA helicase may function in this capacity. nsp13 associates with the RTC (57,58). While incorporating nucleotides, the RTC may overcome helicase action. When stalled, however, the helicase is situated to push the RTC backwards (57,58). We suggest two potential scenarios: one where the replicase backtracks and one where the replicase is displaced backward. In the former, the product strand exits through the NTP channel, revealing the product strand 3'-end for mismatch correction by ExoN (57,58,84). In the latter, the displacement of the replicase leaves the duplex strand annealed and accessible to ExoN for mismatch correction (panel 3 in Figure 11B). In both cases, establishing a tug of war between the polymerase and helicase in the RTC would favor exposure of any 3'-end that impedes forward motion of the polymerase. In such a scenario, access would regulate processing by ExoN and may explain the lack of a strict specificity for a mismatched end. So, what would happen in the presence of a catalytically inactive ExoN? It is likely that the complex would associate with the 3'-end, as illustrated in panel 3 of Figure 11B, and create a physical block to synthesis by the RTC. It would be interesting to determine the extent to which substitutions that impair association of ExoN with the RNA duplex phenocopy those that preclude catalysis.

Remdesivir is one nucleotide analog that shows efficacy against SARS-CoV-2 in humans (1). The mechanism of action clearly relates to perturbed dynamics of the RTC (8,31,33,85). Whether or not some perturbations can oc-

cur immediately or only when present at the $n + 4$ position has not been completely resolved. If remdesivir-terminated RNA were ever present, then our results would suggest that attempts at excision by ExoN would not be as facile as other nucleotide pairs (Figure 7). This particular observation highlights the utility of monitoring hydrolysis by ExoN with single-nucleotide resolution.

The structure of ExoN reveals a complex of nsp10 and nsp14 at a 1:1 stoichiometry (19,22,29,46,47,49). However, such an observation does not infer picomolar affinity. We make this statement because the vast majority of the published studies on SARS-CoV-2 ExoN emphasize the stoichiometry of nsp10:nsp14 used without any consideration of the concentrations of each protein present. Our studies suggest that the dissociation constant for the nsp10–nsp14 complex is in the micromolar range (Figure 4). Because our major conclusions derive from studies of the steady state at saturating concentrations of RNA, knowledge of the precise concentration of ExoN is not essential. However, conclusions from single-turnover experiments would be compromised. Direct studies of the nsp10–nsp14 binding equilibrium (Step 1 in Figure 11A) will be essential to elucidation of the kinetic and chemical mechanisms.

Together, we describe a robust system for evaluating co-transcriptional proofreading of the SARS-CoV-2 polymerase and associated factor by the exoribonuclease. The framework established here will hopefully help the field at large by highlighting some of the key gaps in our understanding of the mechanism of ExoN-catalyzed nucleotide excision. Addressing these gaps will yield a clear picture of the structure–dynamics–function relationship of ExoN that can be used to guide design of ExoN-resistant nucleotide analogs.

DATA AVAILABILITY

All data are incorporated into the article and its online supplementary material. Constructs and data sets presented in this study are available upon request.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

Our thanks to Roel Fleuren and Efra Rivera-Serrano for their contributions to the production of figures and models. We thank Stewart Shuman for providing the expression plasmid for RtcA.

FUNDING

National Institutes of Health [R01AI161841 to C.E.C., J.J.A., D.D.; P01CA234228, R01GM110129 to D.A.H., R01AI158463 to R.K.]; University of Minnesota, Office of the Vice President for Research, Grant-in-Aid (to D.A.H.); Interdisciplinary Center for Clinical Research (IZKF) at the University Hospital of the University of Erlangen-Nuremberg, the German Research Foundation [DFG-DU-1872/4-1 to D.D.]; Netherlands Ministry of Education, Cul-

ture and Science (OCW); Netherlands Organization for Scientific Research (NWO) [024.003.019 to D.D.]. Funding for open access charge: National Institutes of Health.

Conflict of interest statement. None declared.

REFERENCES

1. Beigel, J.H., Tomashek, K.M., Dodd, L.E., Mehta, A.K., Zingman, B.S., Kalil, A.C., Hohmann, E., Chu, H.Y., Luetkemeyer, A., Kline, S. *et al.* (2020) Remdesivir for the treatment of Covid-19 - Final report. *N. Engl. J. Med.*, **383**, 1813–1826.
2. Jayk Bernal, A., Gomes da Silva, M.M., Musungaie, D.B., Kovalchuk, E., Gonzalez, A., Delos Reyes, V., Martin-Quiros, A., Caraco, Y., Williams-Diaz, A., Brown, M.L. *et al.* (2022) Molnupiravir for oral treatment of Covid-19 in nonhospitalized patients. *N. Engl. J. Med.*, **386**, 509–520.
3. Lamb, Y.N. (2022) Nirmatrelvir plus Ritonavir: first approval. *Drugs*, **82**, 585–591.
4. Seley-Radtke, K.L. and Yates, M.K. (2018) The evolution of nucleoside analogue antivirals: a review for chemists and non-chemists. Part 1: early structural modifications to the nucleoside scaffold. *Antiviral Res.*, **154**, 66–86.
5. Johnson, K.A. and Dangerfield, T. (2021) Mechanisms of inhibition of viral RNA replication by nucleotide analogs. *Enzymes*, **49**, 39–62.
6. Maheden, K., Todd, B., Gordon, C.J., Tchesnokov, E.P. and Gotte, M. (2021) Inhibition of viral RNA-dependent RNA polymerases with clinically relevant nucleotide analogs. *Enzymes*, **49**, 315–354.
7. Slusarczyk, M., Serpi, M. and Pertusati, F. (2018) Phosphoramidates and phosphonamidates (ProTides) with antiviral activity. *Antivir. Chem. Chemother.*, **26**, 2040206618775243.
8. Seifert, M., Bera, S.C., van Nies, P., Kirchdoerfer, R.N., Shannon, A., Le, T.T., Meng, X., Xia, H., Wood, J.M., Harris, L.D. *et al.* (2021) Inhibition of SARS-CoV-2 polymerase by nucleotide analogs from a single-molecule perspective. *Elife*, **10**, e70968.
9. Swanstrom, R. and Schinazi, R.F. (2022) Lethal mutagenesis as an antiviral strategy. *Science*, **375**, 497–498.
10. Dulin, D., Arnold, J.J., van Laar, T., Oh, H.S., Lee, C., Perkins, A.L., Harki, D.A., Depken, M., Cameron, C.E. and Dekker, N.H. (2017) Signatures of nucleotide analog incorporation by an RNA-Dependent RNA polymerase revealed using high-Throughput magnetic tweezers. *Cell Rep.*, **21**, 1063–1076.
11. Malone, B., Urakova, N., Snijder, E.J. and Campbell, E.A. (2022) Structures and functions of coronavirus replication-transcription complexes and their relevance for SARS-CoV-2 drug design. *Nat. Rev. Mol. Cell Biol.*, **23**, 21–39.
12. Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D. and Baric, R.S. (2011) Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.*, **8**, 270–279.
13. Robson, F., Khan, K.S., Le, T.K., Paris, C., Demirbag, S., Barfuss, P., Rocchi, P. and Ng, W.L. (2020) Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell*, **79**, 710–727.
14. Ogando, N.S., Ferron, F., Decroly, E., Canard, B., Posthuma, C.C. and Snijder, E.J. (2019) The Curious case of the nidovirus exoribonuclease: its role in RNA synthesis and replication fidelity. *Front. Microbiol.*, **10**, 1813.
15. Minskaia, E., Hertzog, T., Gorbalenya, A.E., Campanacci, V., Cambillau, C., Canard, B. and Ziebuhr, J. (2006) Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Nat. Acad. Sci. U.S.A.*, **103**, 5108–5113.
16. Snijder, E.J., Bredenbeek, P.J., Dobbe, J.C., Thiel, V., Ziebuhr, J., Poon, L.L., Guan, Y., Rozanov, M., Spaan, W.J. and 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.
17. Beese, L.S. and Steitz, T.A. (1991) Structural basis for the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.
18. Chen, P., Jiang, M., Hu, T., Liu, Q., Chen, X.S. and Guo, D. (2007) Biochemical characterization of exoribonuclease encoded by SARS coronavirus. *J. Biochem. Mol. Biol.*, **40**, 649–655.
19. Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R. and Rao, Z. (2015) Structural basis and functional analysis

- of the SARS coronavirus nsp14-nsp10 complex. *Proc. Nat. Acad. Sci. U.S.A.*, **112**, 9436–9441.
20. Bouvet, M., Imbert, I., Subissi, L., Gluais, L., Canard, B. and Decroly, E. (2012) RNA 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex. *Proc. Nat. Acad. Sci. U.S.A.*, **109**, 9372–9377.
 21. Bouvet, M., Lugari, A., Posthuma, C.C., Zevenhoven, J.C., Bernard, S., Betzi, S., Imbert, I., Canard, B., Guillemot, J.C., Lecine, P. *et al.* (2014) Coronavirus Nsp10, a critical co-factor for activation of multiple replicative enzymes. *J. Biol. Chem.*, **289**, 25783–25796.
 22. Czarna, A., Plewka, J., Kresik, L., Matsuda, A., Karim, A., Robinson, C., O'Byrne, S., Cunningham, F., Georgiou, I., Wilk, P. *et al.* (2022) Refolding of lid subdomain of SARS-CoV-2 nsp14 upon nsp10 interaction releases exonuclease activity. *Structure*, **30**, 1050–1054.
 23. Ogando, N.S., Zevenhoven-Dobbe, J.C., van der Meer, Y., Bredenbeek, P.J., Posthuma, C.C. and Snijder, E.J. (2020) The enzymatic activity of the nsp14 exoribonuclease is critical for replication of MERS-CoV and SARS-CoV-2. *J. Virol.*, **94**, e01246-20.
 24. Becares, M., Pascual-Iglesias, A., Nogales, A., Sola, I., Enjuanes, L. and Zuniga, S. (2016) Mutagenesis of coronavirus nsp14 reveals its potential role in modulation of the Innate immune response. *J. Virol.*, **90**, 5399–5414.
 25. Smith, E.C., Blanc, H., Surdel, M.C., Vignuzzi, M. and Denison, M.R. (2013) Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.*, **9**, e1003565.
 26. Deval, J. and Gurard-Levin, Z.A. (2022) Opportunities and challenges in targeting the proofreading activity of SARS-CoV-2 polymerase complex. *Molecules*, **27**, 2918.
 27. Reha-Krantz, L.J. (2010) DNA polymerase proofreading: multiple roles maintain genome stability. *Biochim. Biophys. Acta*, **1804**, 1049–1063.
 28. Carvajal-Maldonado, D., Drogalis Beckham, L., Wood, R.D. and Doublé, S. (2021) When DNA polymerases multitask: functions beyond nucleotidyl transfer. *Front. Mol. Biosci.*, **8**, 815845.
 29. Liu, C., Shi, W., Becker, S.T., Schatz, D.G., Liu, B. and Yang, Y. (2021) Structural basis of mismatch recognition by a SARS-CoV-2 proofreading enzyme. *Science*, **373**, 1142–1146.
 30. Hillen, H.S., Kocic, G., Farnung, L., Dienemann, C., Tegunov, D. and Cramer, P. (2020) Structure of replicating SARS-CoV-2 polymerase. *Nature*, **584**, 154–156.
 31. Gordon, C.J., Lee, H.W., Tchesnokov, E.P., Perry, J.K., Feng, J.Y., Bilello, J.P., Porter, D.P. and Gotte, M. (2022) Efficient incorporation and template-dependent polymerase inhibition are major determinants for the broad-spectrum antiviral activity of remdesivir. *J. Biol. Chem.*, **298**, 101529.
 32. Kabinger, F., Stiller, C., Schmitzova, J., Dienemann, C., Kocic, G., Hillen, H.S., Hobartner, C. and Cramer, P. (2021) Mechanism of molnupiravir-induced SARS-CoV-2 mutagenesis. *Nat. Struct. Mol. Biol.*, **28**, 740–746.
 33. Tchesnokov, E.P., Gordon, C.J., Woolner, E., Kocinkova, D., Perry, J.K., Feng, J.Y., Porter, D.P. and Gotte, M. (2020) Template-dependent inhibition of coronavirus RNA-dependent RNA polymerase by remdesivir reveals a second mechanism of action. *J. Biol. Chem.*, **295**, 16156–16165.
 34. Das, U. and Shuman, S. (2013) 2'-Phosphate cyclase activity of RtcA: a potential rationale for the operon organization of RtcA with an RNA repair ligase RtcB in Escherichia coli and other bacterial taxa. *RNA*, **19**, 1355–1362.
 35. Moorthy, R., Kenedy, S.A., Rodriguez, D.J. and Harki, D.A. (2021) An efficient synthesis of RNA containing GS-441524: the nucleoside precursor of remdesivir. *RSC Adv.*, **11**, 31373–31376.
 36. Passow, K.T., Caldwell, H.S., Ngo, K.A., Arnold, J.J., Antczak, N.M., Narayanan, A., Jose, J., Sturla, S.J., Cameron, C.E., Ciota, A.T. *et al.* (2021) A chemical strategy for intracellular arming of an endogenous broad-spectrum antiviral nucleotide. *J. Med. Chem.*, **64**, 15429–15439.
 37. Arnold, J.J., Bernal, A., Uche, U., Sterner, D.E., Butt, T.R., Cameron, C.E. and Mattern, M.R. (2006) Small ubiquitin-like modifying protein isopeptidase assay based on poliovirus RNA polymerase activity. *Anal. Biochem.*, **350**, 214–221.
 38. Fairhead, M. and Howarth, M. (2015) Site-specific biotinylation of purified proteins using BirA. *Methods Mol. Biol.*, **1266**, 171–184.
 39. Raducanu, V.S., Tehseen, M., Shirbini, A., Raducanu, D.V. and Hamdan, S.M. (2020) Two chromatographic schemes for protein purification involving the biotin/avidin interaction under native conditions. *J. Chromatogr. A*, **1621**, 461051.
 40. Studier, F.W. (2005) Protein production by auto-induction in high density shaking cultures. *Protein Expression Purif.*, **41**, 207–234.
 41. Studier, F.W. (2014) Stable expression clones and auto-induction for protein production in E. coli. *Methods Mol. Biol.*, **1091**, 17–32.
 42. Smidansky, E.D., Arnold, J.J., Reynolds, S.L. and Cameron, C.E. (2011) Human mitochondrial RNA polymerase: evaluation of the single-nucleotide-addition cycle on synthetic RNA/DNA scaffolds. *Biochemistry*, **50**, 5016–5032.
 43. Arnold, J.J., Ghosh, S.K., Bevilacqua, P.C. and Cameron, C.E. (1999) Single-nucleotide resolution of RNA strands in the presence of their RNA complements. *BioTechniques*, **27**, 450–452.
 44. Maio, N., Lafont, B.A.P., Sil, D., Li, Y., Bollinger, J.M. Jr, Krebs, C., Pierson, T.C., Linehan, W.M. and Rouault, T.A. (2021) Fe-S cofactors in the SARS-CoV-2 RNA-dependent RNA polymerase are potential antiviral targets. *Science*, **373**, 236–241.
 45. Baddock, H.T., Brolih, S., Yosaatmadja, Y., Ratnaweera, M., Bielinski, M., Swift, L.P., Cruz-Migoni, A., Fan, H., Keown, J.R., Walker, A.P. *et al.* (2022) Characterization of the SARS-CoV-2 ExoN (nsp14ExoN-nsp10) complex: implications for its role in viral genome stability and inhibitor identification. *Nucleic Acids Res.*, **50**, 1484–1500.
 46. Ferron, F., Subissi, L., Silveira De Moraes, A.T., Le, N.T.T., Sevajol, M., Gluais, L., Decroly, E., Vonnrhein, C., Bricogne, G., Canard, B. *et al.* (2018) Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus RNA. *Proc. Nat. Acad. Sci. U.S.A.*, **115**, E162–E171.
 47. Lin, S., Chen, H., Chen, Z., Yang, F., Ye, F., Zheng, Y., Yang, J., Lin, X., Sun, H., Wang, L. *et al.* (2021) Crystal structure of SARS-CoV-2 nsp10 bound to nsp14-ExoN domain reveals an exoribonuclease with both structural and functional integrity. *Nucleic Acids Res.*, **49**, 5382–5392.
 48. Ma, Z., Pourfarjam, Y. and Kim, I.K. (2021) Reconstitution and functional characterization of SARS-CoV-2 proofreading complex. *Protein Expression Purif.*, **185**, 105894.
 49. Moeller, N.H., Shi, K., Demir, O., Belica, C., Banerjee, S., Yin, L., Durfee, C., Amaro, R.E. and Aihara, H. (2022) Structure and dynamics of SARS-CoV-2 proofreading exoribonuclease ExoN. *Proc. Nat. Acad. Sci. U.S.A.*, **119**, e2106379119.
 50. Riccio, A.A., Sullivan, E.D. and Copeland, W.C. (2022) Activation of the SARS-CoV-2 NSP14 3'-5' exoribonuclease by NSP10 and response to antiviral inhibitors. *J. Biol. Chem.*, **298**, 101518.
 51. Saramago, M., Barria, C., Costa, V.G., Souza, C.S., Viegas, S.C., Domingues, S., Lousa, D., Soares, C.M., Arraiano, C.M. and Matos, R.G. (2021) New targets for drug design: importance of nsp14/nsp10 complex formation for the 3'-5' exoribonucleolytic activity on SARS-CoV-2. *FEBS J.*, **288**, 5130–5147.
 52. Subissi, L., Posthuma, C.C., Collet, A., Zevenhoven-Dobbe, J.C., Gorbalenya, A.E., Decroly, E., Snijder, E.J., Canard, B. and Imbert, I. (2014) One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Nat. Acad. Sci. U.S.A.*, **111**, E3900–E3909.
 53. Deng, X. and Baker, S.C. (2018) An “old” protein with a new story: coronavirus endoribonuclease is important for evading host antiviral defenses. *Virology*, **517**, 157–163.
 54. Ivanov, K.A., Hertzog, T., Rozanov, M., Bayer, S., Thiel, V., Gorbalenya, A.E. and Ziebuhr, J. (2004) Major genetic marker of nidoviruses encodes a replicative endoribonuclease. *Proc. Nat. Acad. Sci. USA*, **101**, 12694–12699.
 55. Hackbart, M., Deng, X. and Baker, S.C. (2020) Coronavirus endoribonuclease targets viral polyuridine sequences to evade activating host sensors. *Proc. Nat. Acad. Sci. U.S.A.*, **117**, 8094–8103.
 56. Kindler, E., Gil-Cruz, C., Spanier, J., Li, Y., Wilhelm, J., Rabouw, H.H., Zust, R., Hwang, M., V'kovski, P., Stalder, H. *et al.* (2017) Early endonuclease-mediated evasion of RNA sensing ensures efficient coronavirus replication. *PLoS Pathog.*, **13**, e1006195.
 57. Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P.M.M., Olinares, P.D.B., Maruthi, K., Eng, E.T., Vatanadslar, H., Chait, B.T. *et al.* (2020) Structural basis for helicase-Polymerase coupling in the SARS-CoV-2 replication-Transcription complex. *Cell*, **182**, 1560–1573.

58. Malone, B., Chen, J., Wang, Q., Llewellyn, E., Choi, Y.J., Olinares, P.D.B., Cao, X., Hernandez, C., Eng, E.T., Chait, B.T. *et al.* (2021) Structural basis for backtracking by the SARS-CoV-2 replication-transcription complex. *Proc. Nat. Acad. Sci. U.S.A.*, **118**, e2106379119.
59. Canal, B., McClure, A.W., Curran, J.F., Wu, M., Ulferts, R., Weissmann, F., Zeng, J., Bertolin, A.P., Milligan, J.C., Basu, S. *et al.* (2021) Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of nsp14/nsp10 exoribonuclease. *Biochem. J.*, **478**, 2445–2464.
60. Arnold, J.J. and Cameron, C.E. (2004) Poliovirus RNA-dependent RNA polymerase (3Dpol): pre-steady-state kinetic analysis of ribonucleotide incorporation in the presence of Mg²⁺. *Biochemistry*, **43**, 5126–5137.
61. Johnson, K.A. (2010) The kinetic and chemical mechanism of high-fidelity DNA polymerases. *Biochim. Biophys. Acta*, **1804**, 1041–1048.
62. Griffiths, A.D., Potter, B.V. and Eperon, I.C. (1987) Stereospecificity of nucleases towards phosphorothioate-substituted RNA: stereochemistry of transcription by T7 RNA polymerase. *Nucleic Acids Res.*, **15**, 4145–4162.
63. Liu, J. and Tsai, M.D. (2001) DNA polymerase beta: pre-steady-state kinetic analyses of dATP alpha S stereoselectivity and alteration of the stereoselectivity by various metal ions and by site-directed mutagenesis. *Biochemistry*, **40**, 9014–9022.
64. Romaniuk, P.J. and Eckstein, F. (1982) A study of the mechanism of T4 DNA polymerase with diastereomeric phosphorothioate analogues of deoxyadenosine triphosphate. *J. Biol. Chem.*, **257**, 7684–7688.
65. Brautigam, C.A. and Steitz, T.A. (1998) Structural principles for the inhibition of the 3'-5' exonuclease activity of Escherichia coli DNA polymerase I by phosphorothioates. *J. Mol. Biol.*, **277**, 363–377.
66. Eckstein, F. (1985) Nucleoside phosphorothioates. *Annu. Rev. Biochem.*, **54**, 367–402.
67. Eckstein, F. (2014) Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Ther.*, **24**, 374–387.
68. Putney, S.D., Benkovic, S.J. and Schimmel, P.R. (1981) A DNA fragment with an alpha-phosphorothioate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated in vivo. *Proc. Nat. Acad. Sci. U.S.A.*, **78**, 7350–7354.
69. Yang, Z., Sismour, A.M. and Benner, S.A. (2007) Nucleoside alpha-thiotriphosphates, polymerases and the exonuclease III analysis of oligonucleotides containing phosphorothioate linkages. *Nucleic Acids Res.*, **35**, 3118–3127.
70. Elfiky, A.A. (2020) Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. *Life Sci.*, **253**, 117592.
71. Lim, Y.S., Nguyen, L.P., Lee, G.H., Lee, S.G., Lyoo, K.S., Kim, B. and Hwang, S.B. (2021) Asunaprevir, a potent Hepatitis C virus protease inhibitor, blocks SARS-CoV-2 propagation. *Mol. Cells*, **44**, 688–695.
72. Lo, H.S., Hui, K.P.Y., Lai, H.M., He, X., Khan, K.S., Kaur, S., Huang, J., Li, Z., Chan, A.K.N., Cheung, H.H. *et al.* (2021) Simeprevir potently suppresses SARS-CoV-2 replication and synergizes with Remdesivir. *ACS Cent Sci*, **7**, 792–802.
73. Gizzi, A.S., Grove, T.L., Arnold, J.J., Jose, J., Jangra, R.K., Garforth, S.J., Du, Q., Cahill, S.M., Dulyaninova, N.G., Love, J.D. *et al.* (2018) A naturally occurring antiviral ribonucleotide encoded by the human genome. *Nature*, **558**, 610–614.
74. Jones, A.N., Mourao, A., Czarna, A., Matsuda, A., Fino, R., Pyrc, K., Sattler, M. and Popowicz, G.M. (2022) Characterization of SARS-CoV-2 replication complex elongation and proofreading activity. *Sci. Rep.*, **12**, 9593.
75. Bera, S.C., Seifert, M., Kirchdoerfer, R.N., van Nies, P., Wubulikasimu, Y., Quack, S., Papini, F.S., Arnold, J.J., Canard, B., Cameron, C.E. *et al.* (2021) The nucleotide addition cycle of the SARS-CoV-2 polymerase. *Cell Rep.*, **36**, 109650.
76. Rivera-Serrano, E.E., Gizzi, A.S., Arnold, J.J., Grove, T.L., Almo, S.C. and Cameron, C.E. (2020) Viperin reveals its true function. *Annu Rev Virol*, **7**, 421–446.
77. Jahns, H., Taneja, N., Willoughby, J.L.S., Akabane-Nakata, M., Brown, C.R., Nguyen, T., Bisbe, A., Matsuda, S., Hettlinger, M., Manoharan, R.M. *et al.* (2022) Chirality matters: stereo-defined phosphorothioate linkages at the termini of small interfering RNAs improve pharmacology in vivo. *Nucleic Acids Res.*, **50**, 1221–1240.
78. Chou, T.F., Baraniak, J., Kaczmarek, R., Zhou, X., Cheng, J., Ghosh, B. and Wagner, C.R. (2007) Phosphoramidate pronucleotides: a comparison of the phosphoramidase substrate specificity of human and Escherichia coli histidine triad nucleotide binding proteins. *Mol Pharm.*, **4**, 208–217.
79. Ozga, M., Dolot, R., Janicka, M., Kaczmarek, R. and Krakowiak, A. (2010) Histidine triad nucleotide-binding protein 1 (HINT-1) phosphoramidase transforms nucleoside 5'-O-phosphorothioates to nucleoside 5'-O-phosphates. *J. Biol. Chem.*, **285**, 40809–40818.
80. Li, Y., Yang, B., Quan, Y. and Li, Z. (2021) Advancement of prodrug approaches for nucleotide antiviral agents. *Curr. Top. Med. Chem.*, **21**, 2909–2927.
81. Serpi, M., Madela, K., Pertusati, F. and Slusarczyk, M. (2013) Synthesis of phosphoramidate prodrugs: proTide approach. *Curr. Protoc. Nucleic Acid Chem.*, **Chapter 15**, Unit15 15.
82. Serpi, M. and Pertusati, F. (2021) An overview of ProTide technology and its implications to drug discovery. *Expert Opin. Drug Discov.*, **16**, 1149–1161.
83. Meier, C. (2017) Nucleoside diphosphate and triphosphate prodrugs - An unsolvable task? *Antivir. Chem. Chemother.*, **25**, 69–82.
84. Yan, L., Ge, J., Zheng, L., Zhang, Y., Gao, Y., Wang, T., Huang, Y., Yang, Y., Gao, S., Li, M. *et al.* (2021) Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell*, **184**, 184–193.
85. Kocic, G., Hillen, H.S., Tegunov, D., Dienemann, C., Seitz, F., Schmitzova, J., Farnung, L., Siewert, A., Hobartner, C. and Cramer, P. (2021) Mechanism of SARS-CoV-2 polymerase stalling by remdesivir. *Nat. Commun.*, **12**, 279.

Nucleotide Analogues as Inhibitors of SARS-CoV-2 Polymerase, a Key Drug Target for COVID-19

Minchen Chien, Thomas K. Anderson, Steffen Jockusch, Chuanjuan Tao, Xiaoxu Li, Shiv Kumar, James J. Russo, Robert N. Kirchdoerfer, and Jingyue Ju*

Cite This: *J. Proteome Res.* 2020, 19, 4690–4697

Read Online

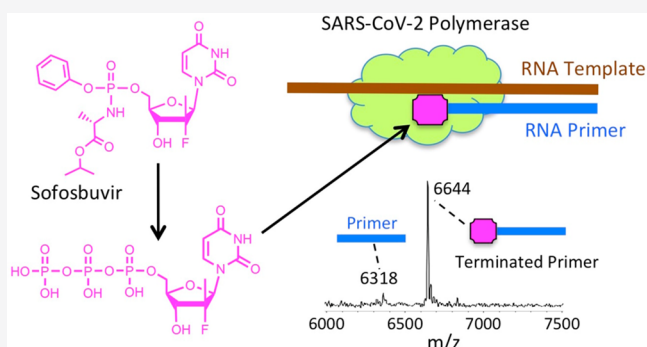
ACCESS |

Metrics & More

Article Recommendations

ABSTRACT: SARS-CoV-2 is responsible for the current COVID-19 pandemic. On the basis of our analysis of hepatitis C virus and coronavirus replication, and the molecular structures and activities of viral inhibitors, we previously demonstrated that three nucleotide analogues (the triphosphates of Sofosbuvir, Alovudine, and AZT) inhibit the SARS-CoV RNA-dependent RNA polymerase (RdRp). We also demonstrated that a library of additional nucleotide analogues terminate RNA synthesis catalyzed by the SARS-CoV-2 RdRp, a well-established drug target for COVID-19. Here, we used polymerase extension experiments to demonstrate that the active triphosphate form of Sofosbuvir (an FDA-approved hepatitis C drug) is incorporated by SARS-CoV-2 RdRp and blocks further incorporation. Using the molecular insight gained from the previous studies, we selected the active triphosphate forms of six other antiviral agents, Alovudine, Tenofovir alafenamide, AZT, Abacavir, Lamivudine, and Emtricitabine, for evaluation as inhibitors of the SARS-CoV-2 RdRp and demonstrated the ability of these viral polymerase inhibitors to be incorporated by SARS-CoV-2 RdRp, where they terminate further polymerase extension with varying efficiency. These results provide a molecular basis for inhibition of the SARS-CoV-2 RdRp by these nucleotide analogues. If sufficient efficacy of some of these FDA-approved drugs in inhibiting viral replication in cell culture is established, they may be explored as potential COVID-19 therapeutics.

KEYWORDS: COVID-19, SARS-CoV-2, RNA-dependent RNA polymerase, nucleotide analogues



INTRODUCTION

The COVID-19 pandemic, caused by SARS-CoV-2, has already infected more than 14 million people worldwide resulting in over 600 000 reported deaths, with severe social and economic ramifications. SARS-CoV-2 is a new member of the subgenus *Sarbecovirus* in the Orthocoronavirinae subfamily, which also includes MERS-CoV and SARS-CoV.¹ The coronaviruses are single-strand RNA viruses, sharing properties with other single-stranded RNA viruses such as hepatitis C virus (HCV), West Nile virus, Marburg virus, HIV virus, Ebola virus, dengue virus, and rhinoviruses. SARS-CoV-2 is a positive-sense single-strand RNA virus like HCV and other flaviviruses;^{2,3} these viruses share a similar replication mechanism requiring an RNA-dependent RNA polymerase (RdRp).

There are currently no effective FDA-approved drugs to specifically treat coronavirus infections such as SARS, MERS, and now COVID-19. Components of nearly every stage of the coronavirus replication cycle have been targeted for drug development.² In particular, the coronavirus RdRp is a well-established drug target. This polymerase shares similar catalytic

mechanisms and displays active site conservation among different positive-sense RNA viruses, including coronaviruses and HCV.⁴ Like RdRps in other viruses, the coronavirus enzyme is highly error-prone,⁵ which might increase its ability to accept modified nucleotide analogues as substrates. Nucleotide analogues that inhibit polymerases are an important group of antiviral agents.^{6–9}

On the basis of our analysis of hepatitis C virus and coronavirus replication, and the molecular structures and activities of viral inhibitors, we previously proposed Sofosbuvir triphosphate as a candidate inhibitor of the SARS-CoV-2 RdRp.^{10,11} Elfiky used a molecular docking study to predict that Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir may have inhibitory activity against SARS-CoV-2

Special Issue: Proteomics in Pandemic Disease

Received: June 3, 2020

Published: July 21, 2020



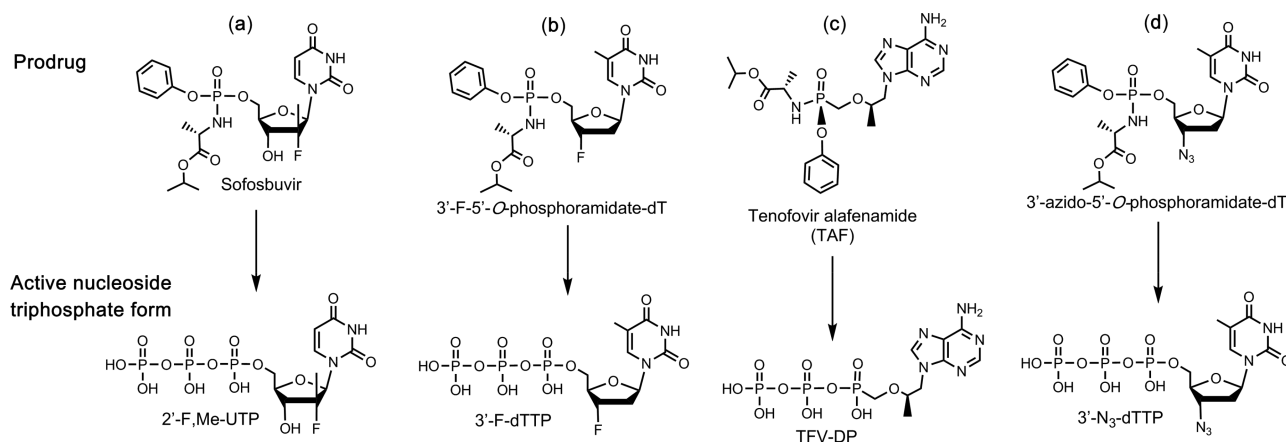


Figure 1. Structures of four prodrug viral inhibitors. Top: Prodrug (phosphoramidate) form; Bottom: Active triphosphate form.

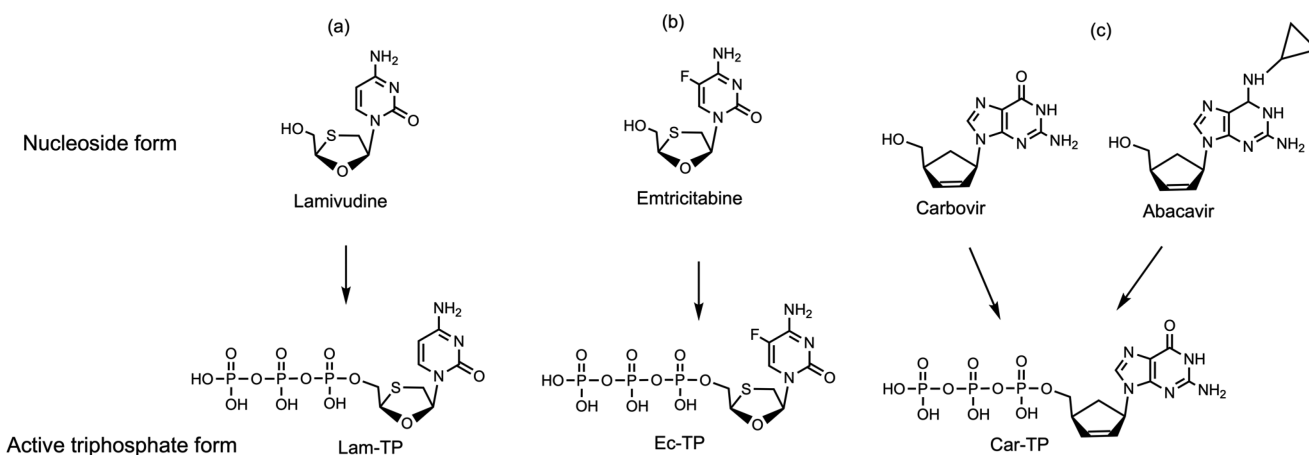


Figure 2. Structures of three viral inhibitors. Top: Nucleoside form; Bottom: Active triphosphate form.

RdRp.¹² Remdesivir, a phosphoramidate prodrug containing a 1'-cyano modification on the sugar, is converted into an adenosine triphosphate analogue inside virus-infected cells, which inhibits the RdRps of MERS-CoV, SARS-CoV, and SARS-CoV-2.^{13,14} Recently, the FDA issued an emergency use authorization for Remdesivir for potential COVID-19 treatment.¹⁵ On the basis of a comparison of the positive-strand RNA genomes of HCV and SARS-CoV-2, Buonaguro et al. postulated that Sofosbuvir might be an optimal nucleotide analogue to repurpose for COVID-19 treatment.¹⁶ After considering the potential advantages of Sofosbuvir, including its low toxicity, its ability to be rapidly activated to the triphosphate form by cellular enzymes, and the high stability of this active molecule intracellularly, Sayad et al. have initiated a clinical trial with Sofosbuvir for treatment of COVID-19.¹⁷ However, a recent kinetic analysis of Sofosbuvir triphosphate with SARS-CoV-2 polymerase indicated that it has lower incorporation activity than UTP.¹⁴

We previously demonstrated that the triphosphates of Sofosbuvir, Alovudine (3'-F-dT), and AZT (3'-N₃-dT) (Figure 1a,b,d) inhibit the SARS-CoV RdRp.¹¹ On the basis of the molecular rationale above, we conducted polymerase primer extension experiments with Sofosbuvir triphosphate (2'-F,Me-UTP, Figure 1a) and demonstrated that it was incorporated by SARS-CoV RdRp and blocked further incorporation. Using the same molecular insight, we selected two HIV reverse transcriptase (RT) inhibitors, Alovudine and AZT, for

evaluation as inhibitors of SARS-CoV RdRp. Alovudine and AZT share a similar backbone structure (base and ribose) with Sofosbuvir but have fewer modification sites (Figure 1b,d). Furthermore, because these modifications on Alovudine and AZT are on the 3' position of the sugar ring in place of the 3'-OH group, if they are accepted as substrates by the RdRp, they will prevent further incorporation of nucleotides leading to obligate termination of RNA synthesis. We demonstrated the ability of the active triphosphate forms of Alovudine and AZT, 3'-F-dTTP (Figure 1b) and 3'-N₃-dTTP (Figure 1d), respectively, to be incorporated by SARS-CoV RdRp where they also terminated further polymerase extension.¹¹ We also demonstrated that a library of additional nucleotide analogues terminate RNA synthesis catalyzed by the SARS-CoV-2 RdRp.¹⁸

We first constructed SARS-CoV-2 RdRp using a similar procedure to that of SARS-CoV,^{19,20} and then we demonstrated that the above three nucleotide analogues (Figure 1a,b,d) are inhibitors of SARS-CoV-2 RdRp. Using structure-activity-based molecular insight, we selected the active triphosphate form of Tenofovir alafenamide (TAF, Vemlidy, an acyclic adenosine nucleotide) (Figure 1c), which is an FDA approved drug for the treatment of HIV and hepatitis B virus (HBV) infection, for evaluation as a SARS-CoV-2 RdRp inhibitor. Similarly, we also selected the triphosphates of three HIV RT inhibitors, Lamivudine triphosphate (Lam-TP, Figure 2a), Emtricitabine triphosphate (Ec-TP, Figure 2b) and

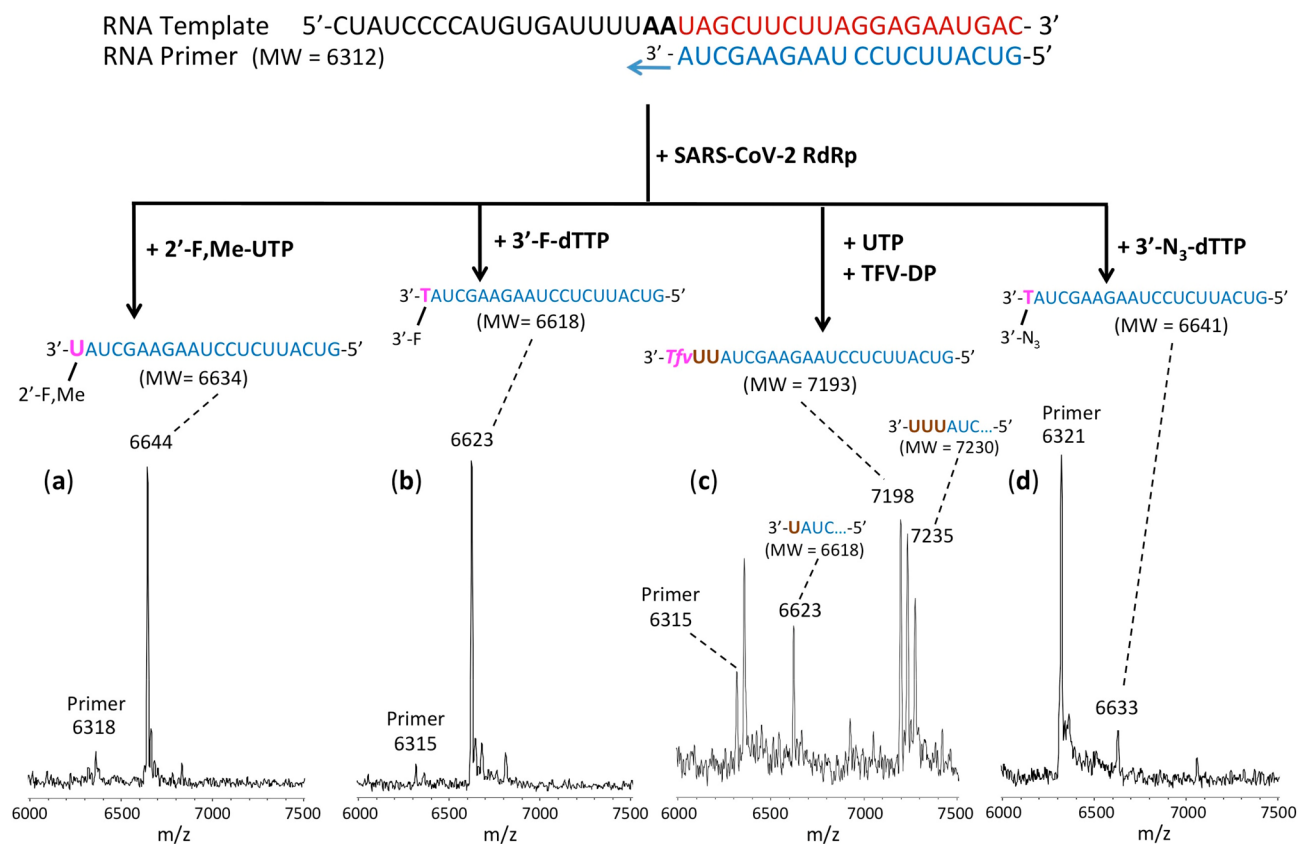


Figure 3. Incorporation of 2'-F,Me-UTP, 3'-F-dTTP, TFV-DP, and 3'-N₃-dTTP by SARS-CoV-2 RdRp to terminate the polymerase reaction. The sequences of the primer and template used for these extension reactions, which are at the 3' end of the SARS-CoV-2 genome, are shown at the top of the figure. Polymerase extension reactions were performed by incubating (a) 2'-F,Me-UTP, (b) 3'-F-dTTP, (c) UTP + TFV-DP, and (d) 3'-N₃-dTTP with preassembled SARS-CoV-2 polymerase (nsp12, nsp7, and nsp8), the indicated RNA template and primer, and the appropriate reaction buffer, followed by detection of reaction products by MALDI-TOF MS. The accuracy for *m/z* determination is ± 10 Da.

Carbovir triphosphate (Car-TP, Figure 2c) to test their ability to inhibit the SARS-CoV-2 RdRp. The results indicated that the active triphosphate forms of Tenofovir, Lamivudine, Emtricitabine, and Abacavir (the prodrug of Car-TP) inhibited this polymerase with varying efficiency. The properties of these four viral inhibitors are described below.

TAF, a prodrug form of the nucleotide analogue viral polymerase inhibitor Tenofovir (TFV), shows potent activity for HIV and HBV but only limited inhibition of host nuclear and mitochondrial polymerases.^{21,22} It is activated by a series of hydrolases to the deprotected monophosphate form, TFV, and then by two consecutive kinase reactions to the triphosphate form Tenofovir diphosphate (TFV-DP).²³ TFV-DP is an acyclic nucleotide and does not have a 3'-OH group. Remarkably, this molecule is incorporated by both HIV and HBV polymerases, terminating nucleic acid elongation and viral replication.^{21,23} In addition, resistance mutations were rarely seen in patients treated with regimens including TAF.²⁴ In view of the fact that the active triphosphate form of TAF, TFV-DP, is much smaller than natural nucleoside triphosphates, we expect that it can easily fit within the active site of the SARS-CoV-2 RdRp. As a noncyclic nucleotide, TFV-DP lacks a normal sugar ring configuration, and thus, we reasoned that it is unlikely to be recognized by 3'-exonucleases involved in SARS-CoV-2 proofreading processes, decreasing the likelihood of developing resistance to the drug.²⁵

The oral drug Lamivudine (3TC) is a cytidine analogue containing an oxathiolane ring with an unnatural (β)-L-

stereochemical configuration, making it a poor substrate for host DNA polymerases.²⁶ This prodrug, which can be taken orally and has low toxicity, is converted by cellular enzymes, first to a monophosphate, then to the active triphosphate form, Lam-TP. Emtricitabine (Emtriva, FTC) has a similar structure to Lamivudine but with a fluorine at the 5-position of the cytosine.²⁷ Conversion of the prodrug form to the active triphosphate is analogous to the activation mechanism for Lamivudine. Like TAF, 3TC and FTC are effective against HBV.²⁸ The absence of an OH group at the 3' position of both Lam-TP and Ec-TP ensures that once these nucleotide analogues are incorporated into the primer in the polymerase reaction, no further incorporation of nucleotides by the polymerase can occur. Car-TP is a carbocyclic guanosine dideoxy-dideoxynucleotide. The parent prodrug, Abacavir (Ziagen), is an FDA-approved nucleoside RT inhibitor used for HIV/AIDS treatment.^{29,30} We previously studied Car-TP as an inhibitor of the SARS-CoV and SARS-CoV-2 RdRp using a higher concentration than in the current study.¹⁸

EXPERIMENTAL SECTION

Materials

Nucleoside triphosphates and nucleoside triphosphate analogues were purchased from TriLink BioTechnologies (CTP, ATP and UTP), Sierra Bioresearch (2'-F,Me-UTP), Amersham Life Sciences (3'-F-dTTP, 3'-N₃-dTTP), Toronto Research Chemicals (Lamivudine-TP, Emtricitabine-TP), or

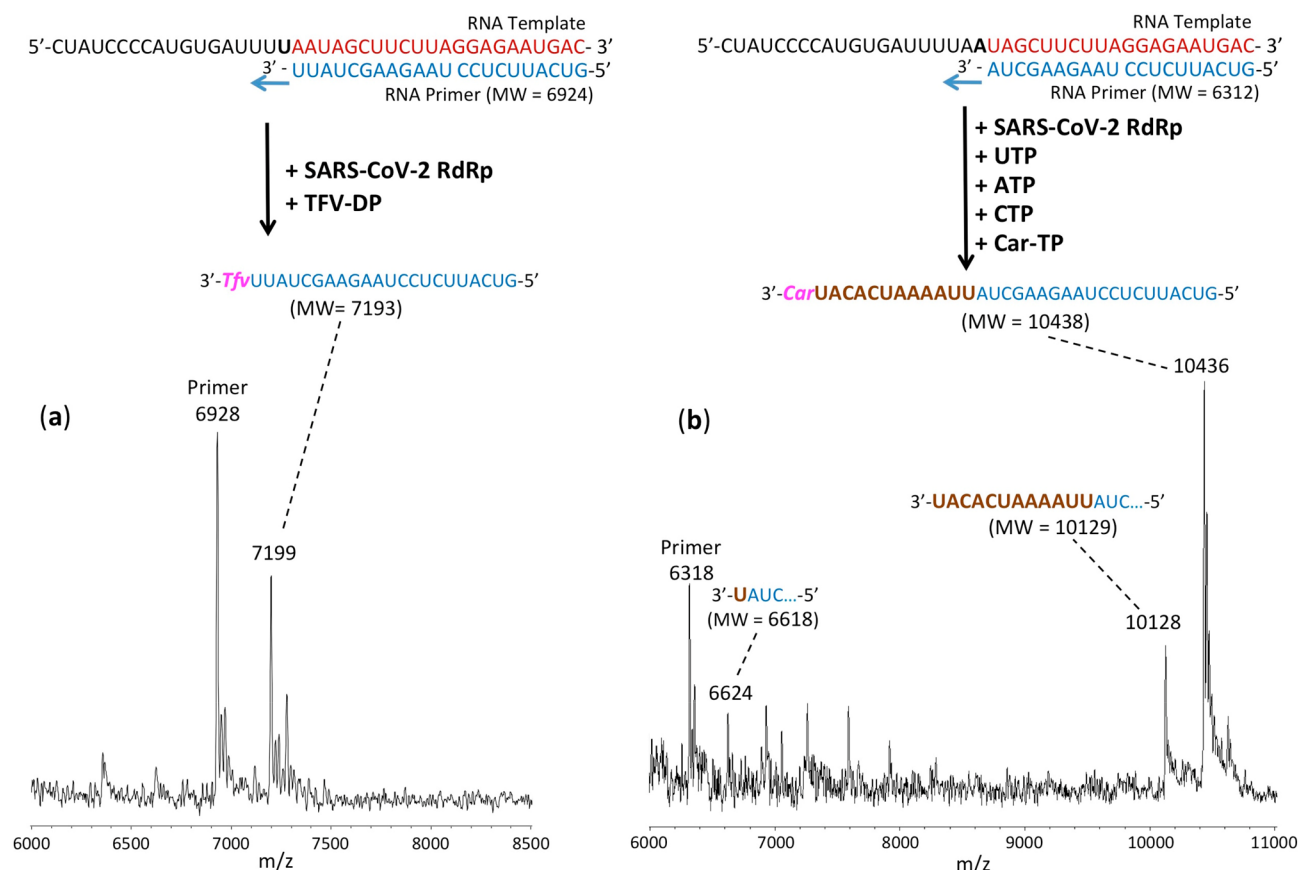


Figure 4. Incorporation of TFV-DP and Car-TP by SARS-CoV-2 RdRp to terminate the polymerase reaction. The sequences of the primers and template used for these extension reactions, which are at the 3' end of the SARS-CoV-2 genome, are shown at the top of the figure. Polymerase extension reactions were performed by incubating (a) TFV-DP and (b) UTP + ATP + CTP + Car-TP with preassembled SARS-CoV-2 polymerase (nsp12, nsp7, and nsp8), the indicated RNA template and primers, and the appropriate reaction buffer, followed by detection of reaction products by MALDI-TOF MS. The accuracy for m/z determination is ± 10 Da.

Santa Cruz Biotechnology (Carbovir-TP). Oligonucleotides were purchased from Integrated DNA Technologies, Inc. or Dharmacon, Inc.

Recombinant Protein Expression of RdRp (nsp12) and Cofactors (nsp7 and nsp8) for SARS-CoV-2

SARS-CoV-2 nsp12. The SARS-CoV-2 nsp12 gene was codon optimized and cloned into pFastBac with C-terminal additions of a thrombin site and double strep tags (Genscript). The pFastBac plasmid and DH10Bac *E. coli* (Life Technologies) were used to create recombinant bacmids. The bacmid was transfected into Sf9 cells (Expression Systems) with Cellfectin II (Life Technologies) to generate recombinant baculovirus. The baculovirus was amplified through two passages in Sf9 cells and then used to infect 1 L of Sf21 cells (Expression Systems) and incubated for 48 h at 27 °C. Cells were harvested by centrifugation and resuspended in wash buffer (25 mM HEPES pH 7.4, 300 mM NaCl, 1 mM MgCl₂, 5 mM DTT) with 143 μ L of BioLock per liter of culture. Cells were lysed via microfluidization (Microfluidics). Lysates were cleared by centrifugation and filtration. The protein was purified using Strep Tactin superflow agarose (IBA). Strep Tactin eluted protein was further purified by size exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Life Sciences) in 25 mM HEPES, 300 mM NaCl, 100 μ M MgCl₂, 2 mM TCEP, at pH 7.4. Pure protein

was concentrated by ultrafiltration prior to flash freezing in liquid nitrogen.

SARS-CoV-2 nsp7 and nsp8. The SARS-CoV-2 nsp7 and nsp8 genes were codon optimized and cloned into pET46 (Novagen) with an N-terminal 6x histidine tag, an enterokinase site, and a TEV protease site. Rosetta2 pLys *E. coli* cells (Novagen) were used for bacterial expression. After induction with isopropyl β -D-1-thiogalactopyranoside (IPTG), cultures were grown at 16 °C for 16 h. Cells were harvested by centrifugation, and pellets were resuspended in wash buffer (10 mM Tris pH 8.0, 300 mM NaCl, 30 mM imidazole, 2 mM DTT). Cells were lysed via microfluidization and lysates were cleared by centrifugation and filtration. Proteins were purified using Ni-NTA agarose beads and eluted with wash buffer containing 300 mM imidazole. Nsp7 and nsp8 proteins were cleaved with 1% (w/w) TEV protease overnight and passed back over Ni-NTA agarose. Cleaved proteins were further purified by size-exclusion chromatography using a Superdex 200 Increase 10/300 column (GE Life Sciences). Purified proteins were concentrated by ultrafiltration prior to flash freezing with liquid nitrogen.

Extension Reactions with SARS-CoV-2 RNA-Dependent RNA Polymerase

The RNA primers and template (sequences shown in Figures 3–5) were annealed by heating to 70 °C for 10 min and cooling to room temperature in 1 \times reaction buffer. For

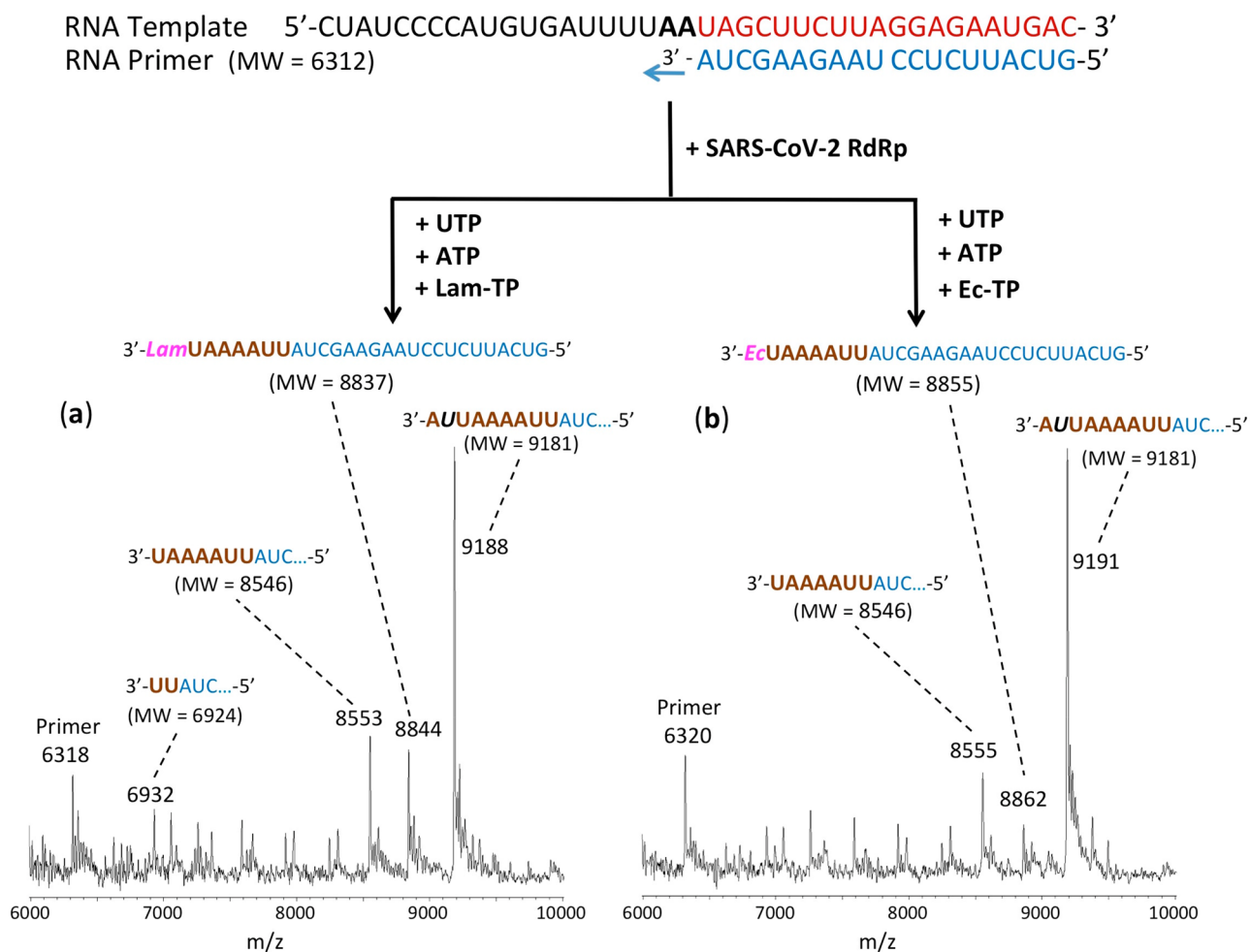


Figure 5. Incorporation of Lam-TP and Ec-TP by SARS-CoV-2 RdRp catalyzed reaction. The sequences of the primer and template used for these extension reactions, which are at the 3' end of the SARS-CoV-2 genome, are shown at the top of the figure. Polymerase extension reactions were performed by incubating (a) UTP + ATP + Lam-TP and (b) UTP + ATP + Ec-TP with preassembled SARS-CoV-2 polymerase (nsp12, nsp7, and nsp8), the indicated RNA template and primer, and the appropriate reaction buffer, followed by detection of reaction products by MALDI-TOF MS. The accuracy for m/z determination is ± 10 Da.

reactions in Figure 3, the RNA polymerase mixtures consisting of 6 μM nsp12 and 18 μM each of cofactors nsp7 and nsp8 were incubated for 15 min at room temperature in a 1:3:3 ratio in 1 \times reaction buffer. For reactions in Figures 4 and 5, higher concentrations of nsp 12, nsp7, and nsp8 were used (10, 30, and 60 μM , respectively). Then 5 μL of the annealed template primer solution containing 2 μM template and 1.7 μM primer in 1 \times reaction buffer was added to 10 μL of the RNA polymerase mixture and incubated for an additional 10 min at room temperature. Finally, 5 μL of a solution containing either 2 mM 2'-F,Me-UTP (Figure 3a), 2 mM 3'-F-dTTP (Figure 3b), 2 mM TFV-DP + 200 μM UTP (Figure 3c), 2 mM 3'-N₃-dTTP (Figure 3d), 2 mM TFV-DP (Figure 4a), 400 μM UTP + 400 μM ATP + 400 μM CTP + 1 mM Car-TP (Figure 4b), 400 μM UTP + 400 μM ATP + 2 mM Lam-TP (Figure 5a) or 400 μM UTP + 400 μM ATP + 2 mM Ec-TP (Figure 5b) in 1 \times reaction buffer was added, and incubation was carried out for 2 h at 30 $^{\circ}\text{C}$. The final concentrations of reagents in the 20 μL extension reactions were 3 μM nsp12, 9 μM nsp7, 9 μM nsp8 (Figure 3) or 5 μM nsp12, 15 μM nsp7, 30 μM nsp8 (Figures 4 and 5), 425 nM RNA primer, 500 nM RNA template, and either 500 μM 2'-F,Me-UTP (Figure 3a), 500 μM 3'-F-dTTP (Figure 3b), 500 μM TFV-DP + 50 μM UTP

(Figure 3c), 500 μM 3'-N₃-dTTP (Figure 3d), 500 μM TFV-DP (Figure 4a), 100 μM UTP + 100 μM ATP + 100 μM CTP + 250 μM Car-TP (Figure 4b), 100 μM UTP + 100 μM ATP + 500 μM Lam-TP (Figure 5a) or 100 μM UTP + 100 μM ATP + 500 μM Ec-TP (Figure 5b). The 1 \times reaction buffer contains the following reagents: 10 mM Tris-HCl pH 8, 10 mM KCl, 2 mM MgCl₂, and 1 mM β -mercaptoethanol. Following desalting using an Oligo Clean & Concentrator (Zymo Research), the samples were subjected to MALDI-TOF-MS (Bruker ultrafleXtreme) analysis.

RESULTS AND DISCUSSION

Given the 98% amino acid similarity of the SARS-CoV and SARS-CoV-2 RdRps and our previous inhibition results on SARS-CoV and SARS-CoV-2 RdRps,^{11,18} we reasoned that the nucleotide analogues listed in Figures 1 and 2 should also inhibit the SARS-CoV-2 polymerase. We thus assessed the ability of 2'-F,Me-UTP, 3'-F-dTTP, TFV-DP, and 3'-N₃-dTTP (the active triphosphate forms of Sofosbuvir, Alovudine, TAF, and AZT, respectively), along with Lam-TP, Ec-TP, and Car-TP (the active triphosphate forms of Lamivudine, Emtricitabine, and Carbvir/Abacavir), to be incorporated by SARS-

CoV-2 RdRp into an RNA primer to terminate the polymerase reaction.

The RdRp of SARS-CoV-2, referred to as nsp12, and its two protein cofactors, nsp7 and nsp8, whose homologues were shown to be required for the processive polymerase activity of nsp12 in SARS-CoV,^{19,20} were cloned and purified as described in the **Experimental Section**. These three viral gene products in SARS-CoV-2 have high homology (e.g., 96% identity and 98% similarity for nsp12, with similar homology levels at the amino acid level for nsp7 and nsp8) to the equivalent gene products from SARS-CoV, the causative agent of SARS.¹¹

We performed polymerase extension assays with 2'-F,Me-UTP, 3'-F-dTTP, 3'-N₃-dTTP, or TFV-DP + UTP, following the addition of a preannealed RNA template and primer to a preassembled mixture of the SARS-CoV-2 RdRp (nsp12) and two cofactor proteins (nsp7 and nsp8). The primer extension products from the reaction were subjected to MALDI-TOF-MS analysis. The RNA template and primer, corresponding to the 3' end of the SARS-CoV-2 genome, were used for the polymerase reaction assay; their sequences are indicated at the top of **Figure 3**. 2'-F,Me-UTP has a 3'-OH group, but because of 2' modification with a fluorine and methyl group, it acts as a nonobligate terminator for HCV RdRp.⁸ 3'-F-dTTP and 3'-N₃-dTTP do not have a 3'-OH, and we previously demonstrated that they are obligate terminators of the SARS-CoV RdRp.¹¹

For the data presented in **Figure 3**, because there are two As in a row in the next available positions of the template for RNA polymerase extension downstream of the priming site, if 2'-F,Me-UTP, 3'-F-dTTP or 3'-N₃-dTTP are incorporated by the viral RdRp and terminate the polymerase reaction, a single nucleotide analogue will be added to the 3'-end of the primer strand. Because the two As in the template are followed by four Us, in the case of the TFV-DP/UTP mixture, two UTPs should be incorporated prior to the incorporation and termination by TFV-DP, which is an ATP analogue and an obligate terminator due to the absence of an OH group. As shown in **Figure 3**, this is exactly what we observed. In the MALDI-TOF MS trace in **Figure 3a**, a peak indicative of the molecular weight of a single nucleotide (2'-F,Me-UMP) primer extension product was obtained (6644 Da observed, 6634 Da expected). Similarly, in the trace in **Figure 3b**, a single extension peak indicative of a single base extension by 3'-F-dTMP is revealed (6623 Da observed, 6618 Da expected), with no further incorporation. In both of the above cases, the primer was nearly completely depleted, indicating that 2'-F,Me-UTP and 3'-F-dTTP are efficient substrates of the RdRp. In the trace in **Figure 3d**, a single extension peak indicative of a single-base extension by 3'-N₃-dTMP is seen (6633 Da observed, 6641 Da expected), with no evidence of further incorporation, though the incorporation efficiency was lower than for 2'-F,Me-UTP and 3'-F-dTTP; further optimization may be required. Finally, in the trace in **Figure 3c**, a peak indicative of the molecular weight of a primer extension product formed by incorporating 2 Us and 1 TFV (an A analogue) is found (7198 Da observed, 7193 Da expected), in addition to other peaks representing partial incorporation (one U, 6623 Da observed, 6618 Da expected) or misincorporation (3 Us, 7235 Da observed, 7230 Da expected). Importantly, once the TFV was incorporated, there was no further extension, indicating it was an obligate terminator for the RdRp. The result of an additional experiment with TFV-DP is

shown in **Figure 4a**, in which a longer RNA primer was used with the same template RNA, allowing direct incorporation of TFV. Again, only a single TFV was incorporated (7199 Da observed, 7193 Da expected), despite the presence of 3 additional Us in the template.

The results for Car-TP, which is a G analogue, are shown in **Figure 4b**. The most prominent extension peak observed indicates extension by UTP, ATP, and CTP followed by complete termination with a Car-TP (10 436 Da observed, 10 438 Da expected). Despite the inclusion of UTP, ATP and CTP in the mixture with Car-TP, no extension past this point was observed, indicating that Car-TP was an obligate terminator of the SARS-CoV-2 RdRp. In addition, some partial extension peaks were seen, e.g., incorporation of one U (6624 Da observed, 6618 Da expected), and extension up to the position just before the first C in the template strand (10 128 Da observed, 10 129 Da expected). These results are consistent with previous results obtained using a higher concentration of Car-TP.¹⁸

MALDI-TOF MS results for extension by the CTP analogues Lam-TP and Ec-TP are shown in panels a and b, respectively, of **Figure 5**. There was relatively poor incorporation by these nucleotide analogues. With Lam-TP, a small peak was observed at 8844 Da (8837 Da expected) indicating the incorporation of Lam-TP following multiple incorporated Us and As. In addition, partial extension peaks were observed at 6932 Da indicating extension by two Us (6924 Da expected) and at 8553 Da indicating extension by 2 Us, 4 As, and 1 U (8546 Da expected). However, the most prominent peak was observed at 9188 Da, indicating misincorporation by a U at the position where the C analogue Lam-TP would be expected to be incorporated followed by incorporation of the subsequent A (9181 Da expected). Similar results were obtained for Ec-TP. Minimal extension by Ec-TP is indicated by the peak at 8862 Da (8855 Da expected), but a partial extension peak indicating incorporation by 2 Us, 4 As and 1 U at 8555 Da (8546 Da expected), and a prominent peak indicating misincorporation by a U at the position where the C analogue Ec-TP should be incorporated and a subsequent A at 9191 Da (9181 Da expected) were also present. These misincorporation results for both Lam-TP and Ec-TP indicate that SARS-CoV-2 RdRp has low fidelity, which is consistent with the known low fidelity of RdRps.⁵

NOTE ADDED IN REVISION

Data for four of the nucleotide analogues included in this paper were presented in a preprint posted on bioRxiv on March 20, 2020.³¹ This field is moving rapidly, and while the current paper was under review and revision, numerous additional publications and preprints have appeared. Sofosbuvir has been shown to inhibit SARS-CoV-2 replication in Huh-2 (human hepatoma-derived) and Calu-3 (Type II pneumocyte-derived) cells with EC50 values of 6.2 and 9.5 μM, respectively, but not in Vero-E6 cells.³² Sofosbuvir was also shown to protect human brain organoids from infection by SARS-CoV-2.³³ A recent preprint provides $K_{1/2}$ values (the concentration leading to 50% SARS-CoV-2 polymerase extension) for a library of nucleotide analogues including Sofosbuvir and others examined in this paper.³⁴ Recently, results from a cohort study comparing COVID-19 outcomes in over 77,000 HIV patients taking combination drugs including Tenofovir and Emtricitabine, among other protocols, indicated that these individuals had a somewhat lower COVID-19 diagnosis rate and suggested

that Tenofovir disoproxil fumarate led to the best overall COVID-19 results.³⁵

CONCLUSIONS

In summary, these results demonstrate that the nucleotide analogues 2'-F,Me-UTP, 3'-F-dTTP, TFV-DP, and Car-TP terminate the RNA synthesis catalyzed by SARS-CoV-2 RdRp. In contrast, 3'-N₃-dTTP, Lam-TP, and Ec-TP were poor RdRp substrates. Sofosbuvir, Tenofovir alafenamide, and Abacavir, the prodrugs of 2'-F,Me-UTP, TFV-DP, and Car-TP, respectively, are FDA-approved oral drugs for treatment of other viral infections, and their safety profiles are well-established. The phosphoramidate prodrugs for Alovudine and Abacavir can be readily synthesized using the ProTide prodrug approach.³⁶ The results presented here, coupled with those we obtained previously,¹⁸ provide a molecular basis for inhibition of the SARS-CoV-2 RdRp by a library of nucleotide analogues. If these FDA-approved drugs display efficacy in inhibiting SARS-CoV-2 replication in cell culture, as recently demonstrated for Sofosbuvir in virus infected lung cells³² and brain organoids,³³ they can be considered as potential candidates in clinical trials for the treatment and prevention of COVID-19.

AUTHOR INFORMATION

Corresponding Author

Jingyue Ju – Center for Genome Technology and Biomolecular Engineering, Departments of Chemical Engineering, and Pharmacology, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-4974-4241; Email: dj222@columbia.edu

Authors

Minchen Chien – Center for Genome Technology and Biomolecular Engineering and Departments of Chemical Engineering, Columbia University, New York, New York 10027, United States

Thomas K. Anderson – Department of Biochemistry and Institute of Molecular Virology, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

Steffen Jockusch – Center for Genome Technology and Biomolecular Engineering and Chemistry, Columbia University, New York, New York 10027, United States; orcid.org/0000-0002-4592-5280

Chuanjuan Tao – Center for Genome Technology and Biomolecular Engineering and Departments of Chemical Engineering, Columbia University, New York, New York 10027, United States

Xiaoxu Li – Center for Genome Technology and Biomolecular Engineering and Departments of Chemical Engineering, Columbia University, New York, New York 10027, United States

Shiv Kumar – Center for Genome Technology and Biomolecular Engineering and Departments of Chemical Engineering, Columbia University, New York, New York 10027, United States

James J. Russo – Center for Genome Technology and Biomolecular Engineering and Departments of Chemical Engineering, Columbia University, New York, New York 10027, United States

Robert N. Kirchdoerfer – Department of Biochemistry and Institute of Molecular Virology, University of Wisconsin-Madison, Madison, Wisconsin 53706, United States

Complete contact information is available at:
<https://pubs.acs.org/10.1021/acs.jproteome.0c00392>

Author Contributions

J.J. and R.N.K. conceived and directed the project; the approaches and assays were designed and conducted by J.J., X.L., S.K., S.J., J.J.R., M.C. and C.T., and SARS-CoV-2 polymerase and associated proteins nsp7 and 8 were cloned and purified by T.K.A. and R.N.K. Data were analyzed by all authors. All authors wrote and reviewed the manuscript.

Funding

This research is supported by Columbia University, a grant from the Jack Ma Foundation, a generous gift from the Columbia Engineering Member of the Board of Visitors Dr. Bing Zhao, and Fast Grants to J.J. and a National Institute of Allergy and Infectious Disease grant AI123498 to R.N.K.

Notes

The authors declare no competing financial interest.


REFERENCES

- (1) Zhu, N.; Zhang, D.; Wang, W.; Li, X.; Yang, B.; Song, J.; Zhao, X.; Huang, B.; Shi, W.; Lu, R.; Niu, P.; Zhan, F.; Ma, X.; Wang, D.; Xu, W.; Wu, G.; Gao, G. F.; Tan, W. for the China Novel Coronavirus Investigating and Research Team. A novel coronavirus from patients with pneumonia in China, 2019. *N. Engl. J. Med.* **2020**, *382*, 727–733.
- (2) Zumla, A.; Chan, J. F. W.; Azhar, E. I.; Hui, D. S. C.; Yuen, K.-Y. Coronaviruses – drug discovery and therapeutic options. *Nat. Rev. Drug Discovery* **2016**, *15*, 327–347.
- (3) Dustin, L. B.; Bartolini, B.; Capobianchi, M. R.; Pistello, M. Hepatitis C virus: life cycle in cells, infection and host response, and analysis of molecular markers influencing the outcome of infection and response to therapy. *Clin. Microbiol. Infect.* **2016**, *22*, 826–832.
- (4) te Velthuis, A. J. W. Common and unique features of viral RNA-dependent polymerases. *Cell. Mol. Life Sci.* **2014**, *71*, 4403–4420.
- (5) Selisko, B.; Papageorgiou, N.; Ferron, F.; Canard, B. Structural and functional basis of the fidelity of nucleotide selection by *Flavivirus* RNA-dependent RNA polymerases. *Viruses* **2018**, *10*, 59.
- (6) McKenna, C. E.; Levy, J. N.; Khawli, L. A.; Harutunian, V.; Ye, T.-G.; Starnes, M. C.; Bapat, A.; Cheng, Y.-C. Inhibitors of viral nucleic acid polymerases. Pyrophosphate analogues. *ACS Symp. Ser.* **1989**, *401*, 1–16.
- (7) Öberg, B. Rational design of polymerase inhibitors as antiviral drugs. *Antiviral Res.* **2006**, *71*, 90–95.
- (8) Eltahla, A. A.; Luciani, F.; White, P. A.; Lloyd, A. R.; Bull, R. A. Inhibitors of the hepatitis C virus polymerase; mode of action and resistance. *Viruses* **2015**, *7*, 5206–5224.
- (9) De Clercq, E.; Li, G. Approved antiviral drugs over the past 50 years. *Clin. Microbiol. Rev.* **2016**, *29*, 695–747.
- (10) Ju, J.; Kumar, S.; Li, X.; Jockusch, S.; Russo, J. J. Nucleotide analogues as inhibitors of viral polymerases. *bioRxiv.* **2020**, DOI: [10.1101/2020.01.30.927574](https://doi.org/10.1101/2020.01.30.927574).
- (11) Ju, J.; Li, X.; Kumar, S.; Jockusch, S.; Chien, M.; Tao, C.; Morozova, I.; Kalachikov, S.; Kirchdoerfer, R. N.; Russo, J. J. Nucleotide analogues as inhibitors of SARS-CoV polymerase. *bioRxiv* **2020**, DOI: [10.1101/2020.03.12.989186](https://doi.org/10.1101/2020.03.12.989186).
- (12) Elfiky, A. A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life Sci.* **2020**, *253*, 117592.
- (13) Gordon, C. J.; Tchesnokov, E. P.; Feng, J. Y.; Porter, D. P.; Götte, M. The antiviral compound remdesivir potently inhibits RNA-dependent RNA polymerase from Middle East respiratory syndrome coronavirus. *J. Biol. Chem.* **2020**, *295*, 4773–4779.

- (14) Gordon, C. J.; Tchesnokov, E. P.; Woolner, E.; Perry, J. K.; Feng, J. Y.; Porter, D. P.; Götte, M. Remdesivir is a direct-acting antiviral that inhibits RNA-dependent RNA polymerase from severe acute respiratory syndrome coronavirus 2 with high potency. *J. Biol. Chem.* **2020**, *295*, 6785–6797.
- (15) Eastman, R. T.; Roth, J. S.; Brimacombe, K. R.; Simeonov, A.; Shen, M.; Patnaik, S.; Hall, M. D. Remdesivir: A review of its discovery and development leading to emergency use authorization for treatment of COVID-19. *ACS Cent. Sci.* **2020**, *6*, 672–683.
- (16) Buonaguro, L.; Tagliamonte, M.; Tornesello, M. L.; Buonaguro, F. M. SARS-CoV-2 RNA polymerase as target for antiviral therapy. *J. Transl. Med.* **2020**, *18*, 185.
- (17) Sayad, B.; Sobhani, M.; Khodarahmi, R. Sofosbuvir as repurposed antiviral drug against COVID-19: why were we convinced to evaluate the drug in a registered/approved clinical trial? *Arch. Med. Res.* **2020**, DOI: 10.1016/j.arcmed.2020.04.018.
- (18) Jockusch, S.; Tao, C.; Li, X.; Anderson, T. K.; Chien, M.; Kumar, S.; Russo, J. J.; Kirchdoerfer, R. N.; Ju, J. A library of nucleotide analogues terminate RNA synthesis catalyzed by polymerases of coronaviruses that cause SARS and COVID-19. *Antiviral Res.* **2020**, *180*, 104857.
- (19) Subissi, L.; Posthuma, C. C.; Collet, A.; Zevenhoven-Dobbe, J. C.; Gorbalenya, A. E.; Decroly, E.; Snijder, E. J.; Canard, B.; Imbert, I. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, E3900–E3909.
- (20) Kirchdoerfer, R. N.; Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* **2019**, *10*, 2342.
- (21) Lou, L. Advances in nucleotide antiviral development from scientific discovery to clinical applications: Tenofovir disoproxil fumarate for hepatitis B. *J. Clin. Translat. Hepatol.* **2013**, *1*, 33–38.
- (22) De Clercq, E. Tenofovir alafenamide (TAF) as the successor of tenofovir disoproxil fumarate (TDF). *Biochem. Pharmacol.* **2016**, *119*, 1–7.
- (23) Birkus, G.; Bam, R. A.; Willkom, M.; Frey, C. R.; Tsai, L.; Stray, K. M.; Yant, S. R.; Cihlar, T. Intracellular activation of tenofovir alafenamide and the effect of viral and host protease inhibitors. *Antimicrob. Agents Chemother.* **2016**, *60*, 316–322.
- (24) Margot, N.; Cox, S.; Das, M.; McCallister, S.; Miller, M. D.; Callebaut, C. Rare emergence of drug resistance in HIV-1 treatment-naïve patients receiving elvitegravir/cobicistat/emtricitabine/tenofovir alafenamide for 144 weeks. *J. Clin. Virol.* **2018**, *103*, 37–42.
- (25) Smith, E. C.; Blanc, H.; Vignuzzi, M.; Denison, M. R. Coronaviruses lacking exoribonuclease activity are susceptible to lethal mutagenesis: evidence for proofreading and potential therapeutics. *PLoS Pathog.* **2013**, *9*, e1003565.
- (26) Quercia, R.; Perno, C.-F.; Koteff, J.; Moore, K.; McCoig, C.; St. Clair, M.; Kuritzkes, D. Twenty-five years of lamivudine: current and future use for the treatment of HIV-1 infection. *JAIDS, J. Acquired Immune Defic. Syndr.* **2018**, *78*, 125–135.
- (27) Hung, M.; Tokarsky, E. J.; Lagpacan, L.; Zhang, L.; Suo, Z.; Lansdon, E. B. Elucidating molecular interactions of L-nucleotides with HIV-1 reverse transcriptase and mechanism of M184V-caused drug resistance. *Commun. Biol.* **2019**, *2*, 469.
- (28) Lim, S. G.; Ng, T. M.; Kung, N.; Krastev, Z.; Volfova, M.; Husa, P.; Lee, S. S.; Chan, S.; Shiffman, M. L.; Washington, M. K.; Rigney, A.; Anderson, J.; Mondou, E.; Snow, A.; Sorbel, J.; Guan, R.; Rousseau, F. A double-blind placebo-controlled study of emtricitabine in chronic hepatitis B. *Arch. Intern. Med.* **2006**, *166*, 49–56.
- (29) Faletto, M. B.; Miller, W. H.; Garvey, E. P.; St. Clair, M. H.; Daluge, S. M.; Good, S. S. Unique intracellular activation of the potent anti-human immunodeficiency virus agent 1592U89. *Antimicrob. Agents Chemother.* **1997**, *41*, 1099–1107.
- (30) Ray, A. S.; Basavapathruni, A.; Anderson, K. S. Mechanistic studies to understand the progressive development of resistance in human immunodeficiency virus type 1 reverse transcriptase to abacavir. *J. Biol. Chem.* **2002**, *277*, 40479–40490.
- (31) Chien, M.; Anderson, T. K.; Jockusch, S.; Tao, C.; Kumar, S.; Li, X.; Russo, J. J.; Kirchdoerfer, R. N.; Ju, J. Nucleotide analogues as inhibitors of SARS-CoV-2 polymerase. *bioRxiv.* **2020**. DOI: 10.1101/2020/03/18.997585.
- (32) Sacramento, C. Q.; Fintelman-Rodrigues, N.; Temerozo, J. R.; da Silva Gomes Dias, S.; Ferreira, A. C.; Mattos, M.; Pão, C. R. R.; de Freitas, C. S.; Cardoso Soares, V.; Bozza, F. A.; Bou-Habib, D. C.; Bozza, P. T.; Souza, T. M. L. The *in vitro* antiviral activity of the anti-hepatitis C virus (HCV) drugs daclatasvir and sofosbuvir against SARS-CoV-2. *bioRxiv.* **2020**, DOI: 10.1101/2020.06.15.153411.
- (33) Mesci, P.; Macia, A.; Saleh, A.; Martin-Sancho, L.; Yin, X.; Sneathlge, C.; Avansini, S.; Chanda, S. K.; Muotri, A. Sofosbuvir protects human brain organoids against SARS-CoV-2. *bioRxiv.* **2020**, DOI: 10.1101/2020.05.30.125856.
- (34) Lu, G.; Zhang, X.; Zheng, W.; Sun, J.; Hua, L.; Xu, L.; Chu, X.; Ding, S.; Xiong, W. Development of a simple *in vitro* assay to identify and evaluate nucleotide analogs against SARS-CoV-2 RNA-dependent RNA polymerase. *bioRxiv.* **2020**, DOI: 10.1101/2020.07.16.205799.
- (35) del Amo, J.; Polo, R.; Moreno, S.; Díaz, A.; Martínez, E.; Arribas, J. R.; Jarrin, I.; Hernán, M. A. The Spanish HIV/COVID-19 Collaboration. Incidence and severity of COVID-19 in HIV-positive persons receiving antiretroviral therapy. A cohort study. *Ann. Intern. Med.* **2020**, DOI: 10.7326/M20-3689.
- (36) Alanazi, A. S.; James, E.; Mehellou, Y. The ProTide prodrug technology: where next? *ACS Med. Chem. Lett.* **2019**, *10*, 2–5.



^1H , ^{13}C , and ^{15}N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7

Marco Tonelli¹ · Chad Rienstra^{1,2} · Thomas K. Anderson^{1,3} · Rob Kirchdoerfer^{1,3} · Katherine Henzler-Wildman^{1,2} 

Received: 3 September 2020 / Accepted: 12 November 2020 / Published online: 20 November 2020
© Springer Nature B.V. 2020

Abstract

The SARS-CoV-2 genome encodes for approximately 30 proteins. Within the international project covid19-nmr, we distribute the spectroscopic analysis of the viral proteins and RNA. Here, we report NMR chemical shift assignments for the protein nsp7. The 83 amino acid nsp7 protein is an essential cofactor in the RNA-dependent RNA polymerase. The polymerase activity and processivity of nsp12 are greatly enhanced by binding 1 copy of nsp7 and 2 copies of nsp8 to form a 160 kD complex. A separate hexadecameric complex of nsp7 and nsp8 (8 copies of each) forms a large ring-like structure. Thus, nsp7 is an important component of several large protein complexes that are required for replication of the large and complex coronavirus genome. We here report the near-complete NMR backbone and sidechain resonance assignment (^1H , ^{13}C , ^{15}N) of isolated nsp7 from SARS-CoV-2 in solution. Further, we derive the secondary structure and compare it to the previously reported assignments and structure of the SARS-CoV nsp7.

Keywords SARS-CoV-2 · Non-structural protein · Solution NMR-spectroscopy · COVID19-NMR

Biological context

SARS-CoV-2, the causative agent of COVID-19, emerged late in 2019 to cause a global pandemic. This novel coronavirus is highly related to SARS-CoV that emerged in 2002. While several coronaviruses routinely infect humans causing mild respiratory symptoms (van der Hoek 2007), SARS-CoV-2 is capable of causing severe respiratory symptoms or even death (Bchetnia et al. 2020). Coronaviruses are enveloped viruses with large positive-sense RNA genomes. At 30 kb or even greater, some of the genomes of the *Nidovirales* order, which includes coronaviruses, are among the largest RNA genomes known, requiring a highly processive RNA synthesis process for their replication.

Upon entering host cells, the coronavirus RNA genome acts as an mRNA to be translated by host ribosomes to produce the viral polyproteins pp1a and pp1ab. These polyproteins are cleaved to produce a suite of 16 non-structural proteins (nsp) that are responsible for replication and transcription of the viral RNA genome (Snijder et al. 2016). The viral nsp assemble membrane enclosed compartments containing the virus RNA-dependent RNA-polymerase. The minimal polymerase complex required for activity in vitro is composed of nsp7, nsp8 and nsp12. While the polymerase active site is contained wholly within nsp12, nsp7 and nsp8 act as essential co-factors for enzyme activity enabling processive RNA synthesis (Subissi et al. 2014). Nsp7 has also been proposed to act with nsp8 as a part of an RNA primase to generate RNA primers for viral RNA synthesis (Imbert et al. 2006; te Velthuis et al. 2012). However, the activity of the primase or its proposed additional role to extend RNA primer has not been universally reproduced (Subissi et al. 2014).

Previous structures of nsp7 show the protein to be composed of four helical regions where the positioning of the N- and C-terminal helical regions are altered upon binding to nsp8 (Johnson et al. 2010; Zhai et al. 2005). Structural determination of the polymerase complexes of SARS-CoV (Kirchdoerfer and Ward 2019) and SARS-CoV-2 (Gao et al.

✉ Katherine Henzler-Wildman
henzlerwildm@wisc.edu

¹ National Magnetic Resonance Facility at Madison (NMRFAM), University of Wisconsin at Madison, Madison, WI 53706, USA

² Department of Biochemistry, University of Wisconsin at Madison, Madison, WI 53706, USA

³ Institute for Molecular Virology, University of Wisconsin at Madison, Madison, WI 53706, USA

2020), show that two subunits of nsp8 bind to nsp12 and that one of these nsp8 subunits interacts with nsp7 giving a 1:2:1 nsp7:nsp8:nsp12 stoichiometry. A separate crystal structure of SARS-CoV nsp7 bound to nsp8, while resembling the nsp7-nsp8 interactions in the nsp7-nsp8-nsp12 cryoEM structures, showed the assembly of a large 8:8 protein complex (Zhai et al. 2005). Lacking solution evidence for this large complex and alternate assemblies for nsp7 and nsp8 observed in crystal structures of feline coronavirus (Xiao et al. 2012) and SARS-CoV-2 (Konkolova et al. 2020) leaves ambiguity as to the biological role of these large nsp7-nsp8 complexes in the virus life cycle. As a well-conserved component of the virus replication machinery, a greater understanding of nsp7 structure and dynamics will accelerate our understanding of this essential protein complex improving models of protein-protein interactions and laying an important foundation for the development of antiviral therapeutics.

Methods and experiments

Construct design

This study uses the SARS-CoV-2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (Wu et al. 2020). This sequence was inserted into a pET46 vector, containing an N-terminal His₆-tag, Ek protease and a tobacco etch virus (TEV) protease cleavage sites. Due to the nature of the TEV protease cleavage site, the purified protein contained one artificial N-terminal residue (G0) preceding the native protein sequence.

Sample preparation

Uniformly ¹³C, ¹⁵N-labelled Nsp7 protein was expressed in *Escherichia coli* strain Rosetta2 pLysS in M9 minimal medium containing 1 g/L ¹⁵NH₄Cl (Cambridge Isotope Laboratories), 4 g/L ¹³C₆-D-glucose (Cambridge Isotope Laboratories) and 100 µg/mL ampicillin. Bacterial cultures were grown to an O.D. 600_{nm} of 0.8 at 37 °C and induced with 0.5 mM IPTG for 14–16 h at 16 °C. The cell pellet was resuspended in buffer A (10 mM HEPES, pH 7.4, 300 mM NaCl, 30 mM imidazole and 2 mM dithiothreitol). The cells were lysed using a microfluidizer operating at 20,000 psi. The lysate was clarified by centrifugation at 25,000 × g for 30 min and then filtration using a 0.45 µm vacuum filter. Clarified supernatant was bound to Ni-NTA agarose (Qia-gen), washed with buffer A and then eluted with buffer A containing 300 mM total imidazole. Protein containing fractions were pooled and cleaved with 1% (w/w) TEV protease over night at room temperature while dialyzing against 10 mM MOPS, pH 7.0, 150 mM NaCl, 2 mM dithiothreitol. TEV protease and tag were removed via a second IMAC

purification. Protein was further purified with a Superdex200 column (GE Life Sciences) using a buffer containing 10 mM MOPS, pH 7.0, 150 mM NaCl, 2 mM dithiothreitol. Fractions containing the purified proteins were concentrated using Amicon Ultra concentrators (Millipore Sigma). The final NMR sample contained 1.7 mM ¹³C, ¹⁵N-nsp7, 10 mM MOPS, pH 7.0, 150 mM NaCl, 2 mM DTT, 0.025% NaN₃, 7% D₂O.

NMR experiments

All experiments for the backbone and side chain assignments of nsp7 were recorded at 298 K using 600 MHz Varian VNMRs and Bruker Avance III spectrometers, equipped with an H/C/N Cryoprobe. All spectra were acquired using standard pulse sequences optimized to achieve the best performance on cryogenic probes and with non-uniform sampling. The set of NMR experiments used for resonance assignments is summarized in Table 1. Proton resonances were calibrated with respect to the signal of 2,2-dimethylsilapentane-5-sulfonic acid (DSS). Nitrogen and carbon chemical shifts were referenced indirectly to the ¹H standard using a conversion factor derived from the ratio of NMR frequencies (Wishart et al. 1995). Spectra were processed using NMRPipe (Delaglio et al. 1995) with SMILE (Ying et al. 2017) for NUS reconstruction and analyzed using NMRFAM-Sparky (Lee et al. 2015).

Assignments and data deposition

The ¹H-¹⁵N HSQC spectrum of nsp7 shows well-dispersed amide signals (Fig. 1). Assignments were performed with i-PINE (Lee et al. 2019) using PINE-Sparky2 automated (Lee and Markley 2018) interface and manual confirmation in NMRFAM-Sparky (Lee et al. 2015). Backbone assignments are 96% complete with G0, S1, and D67 not visible in the ¹⁵N-HSQC spectrum. For the nsp7 sequence (S1-Q83), assignments are 99% complete for Cα, Cβ, and CO (only V66 unassigned), and 98% complete for H^N and N (D67 and S1 unassigned). Secondary structure prediction was performed using chemical shift assignments of five atoms (H^N, Cα, Cβ, CO, N) for a given residue in the sequence with TALOS-N (Shen and Bax 2013). The results for nsp7 are shown in Fig. 2.

With the exception of G0, the aliphatic and aromatic side chain C-H groups for all residues were assigned (> 99% completeness overall, 100% for nsp7 sequence). In addition, the Nδ2-Hδ2 groups of N residues, the Ne2-He2 groups of Q residues, the Ne-He groups of R residues and the Ne1-He1 group of W29 were also assigned. The Nζ-Hζ groups of K residues and the Nδ1-Hδ1 and Ne2-He2 groups of H36 were not assigned.

Table 1 List of experiments collected to perform the sequence specific assignment of nsp7

Experiments	Time Domain Data Size (complex points)				Spectral width (ppm)			ns	Delay time (s)
	t_1	max t_2	max t_3	NUS points	F_1	F_2	F_3		
^1H - ^{15}N N-HSQC	1024	192			20.0 (^1H)	29.6 (^{15}N)		16	1
HNCACB	1024	60	52	1116	20.0 (^1H)	66.3 (^{13}C)	29.6 (^{15}N)	32	1
CBCA(CO)NH	1024	76	52	1400	20.0 (^1H)	66.3 (^{13}C)	29.6 (^{15}N)	32	1
HNCO	1024	52	52	961	20.0 (^1H)	9.9 (^{13}C)	29.6 (^{15}N)	16	1
HN(CA)CO	1024	52	52	961	20.0 (^1H)	9.9 (^{13}C)	29.6 (^{15}N)	48	1
^1H - ^{13}C -HSQC aliphatic	1024	268			20.0 (^1H)	66.3 (^{13}C)		48	1
C(CO)NH	1024	64	48	999	20.0 (^1H)	66.3 (^{13}C)	29.6 (^{15}N)	48	1
HBHA(CO)NH	1024	64	48	1064	20.0 (^1H)	5.3 (^1H)	29.6 (^{15}N)	32	1
H(CCO)NH	1024	64	48	1064	20.0 (^1H)	6.3 (^1H)	29.6 (^{15}N)	64	1
H(C)CH-TOCSY aliphatic	1024	64	76	1536	20.0 (^1H)	8.3 (^1H)	72.9 (^{13}C)	32	1
NOESY ^1H - ^{13}C -HSQC aliphatic	1024	64	64	1513	20.0 (^1H)	12.7 (^1H)	72.9 (^{13}C)	32	1
^1H - ^{13}C -HSQC aromatic	1024	34			16.7 (^1H)	33.1 (^{13}C)		128	1
(HB)CB(CGCD)HE	1024	48			16.7 (^1H)	26.5 (^{13}C)		256	1
H(C)CH-TOCSY aromatic	1024	34	48	512	16.7 (^1H)	3.0 (^1H)	33.1 (^{13}C)	64	1
NOESY ^1H - ^{13}C -HSQC aromatic	1024	64	48	1131	20.0 (^1H)	12.7 (^1H)	33.1 (^{13}C)	32	1

Fig. 1 Assigned ^1H , ^{15}N N-HSQC spectrum of the ^{13}C , ^{15}N -labelled SARS-CoV-2 nsp7 at 1.7 mM concentration in 10 mM MOPS, pH 7.0, 150 mM NaCl, 2 mM DTT, 0.025% NaN_3 and 7% D_2O measured at 298 K on a 600 MHz Agilent NMR Spectrometer with backbone NH chemical shift assignments shown. The inset shows the central region of the spectrum enlarged for clarity

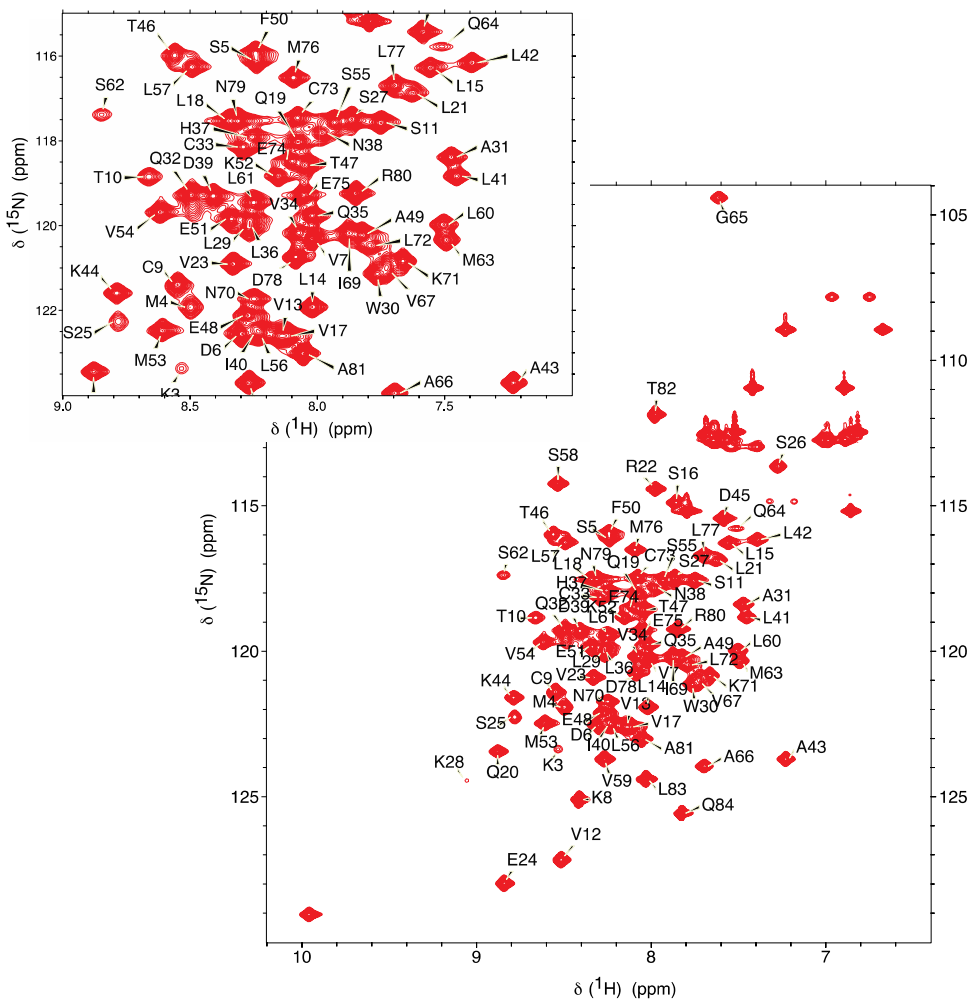
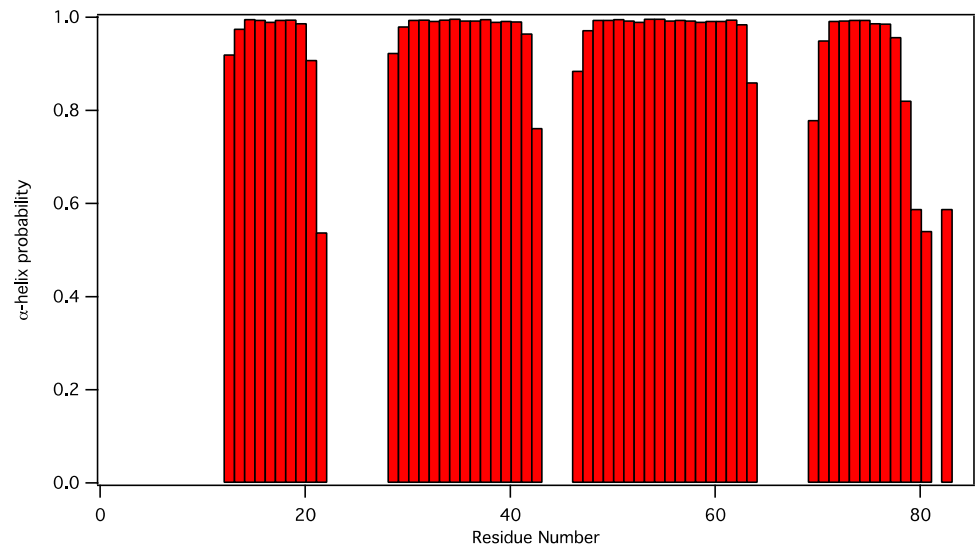


Fig. 2 Display of TALOS-N predicted secondary structure for nsp7. Helical probability shown in red, residues that are highly dynamic, predicted to be coil, or for which there is not a consistent prediction are not shown



The structure of nsp7 from SARS coronavirus was previously determined by NMR by the Wüthrich lab, first at pH 7.5 and high ionic strength (Peti et al. 2005) (BMRB ID 6513, PDB ID 1YSY) and later at pH 6.5 (Johnson et al. 2010) (BMRB ID 16,981, PDB ID 2KYS). The sequence of nsp7 from SARS-2 and SARS are nearly identical, with only a single conservative amino acid difference at position 70 (Fig. 3). The backbone chemical shifts for nsp7 from SARS-2 and SARS are very similar, as might be expected given the very high sequence identity. The dihedral angles predicted by TALOS-N for nsp7 from SARS

are in good agreement with the previous SARS coronavirus nsp7 NMR structure (Fig. 3).

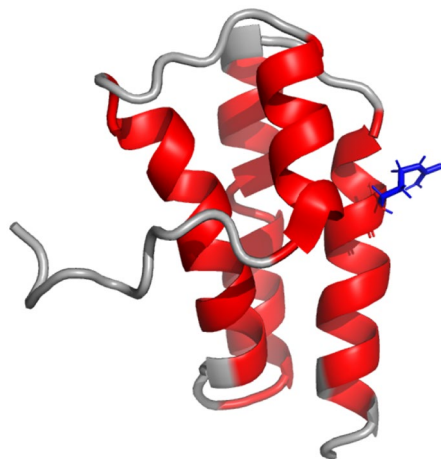
The chemical shift values for the ^1H , ^{13}C and ^{15}N resonances of nsp7 have been deposited at the BioMagResBank (<https://www.bmrwisc.edu>) under accession number 50337. Raw data has been deposited in BMRbig (<https://bmrbig.org/>) under deposition ID bmrbig4.

Fig. 3 Comparison of nsp7 from SARS-CoV-2 and SARS-CoV. **a** Sequence comparison shows a single conservative amino acid substitution. **b** The SARS-CoV-2 helical regions (red) shown in Fig. 2 are plotted on the SARS-CoV nsp7 structure determined by NMR at pH 6.5 (PDB 2KYS). This structure is higher quality than the pH 7.5 structure (PDB 1YSY) because more complete assignments and a larger number of restraints were obtained at pH 6.5. The location of the single amino acid difference between SARS-CoV and SARS-CoV-2 nsp7 is highlighted in blue sticks

(A)

SARS-CoV-2	SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEA
SARS-CoV	SKMSDVKCTSVVLLSVLQQLRVESSSKLWAQCVQLHNDILLAKDTTEA
SARS-CoV-2	FEKMSVLLSVLLSMQGAVDINKLCEEMLDNRATLQ 83
SARS-CoV	FEKMSVLLSVLLSMQGAVDINRLCEEMLDNRATLQ 83

(B)



Acknowledgements Funding for this project was provided by the National Science Foundation EAGER MCB-2031269 and the National Institutes for Health, National Institute for Allergy and Infectious Disease AI123498. This study made use of the National Magnetic Resonance Facility at Madison, which is supported by NIH grant P41GM103399 (NIGMS), old number: P41RR002301. Equipment was purchased with funds from the University of Wisconsin-Madison, the NIH P41GM103399, S10RR02781, S10RR08438, S10RR023438, S10RR025062, S10RR029220), the NSF (DMB-8415048, OIA-9977486, BIR-9214394), and the USDA.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Bchetnia M, Girard C, Duchaine C, Laprise C (2020) The outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2): a review of the current global status. *J Infect Public Health*. <https://doi.org/10.1016/j.jiph.2020.07.011>
- Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277–293
- Gao Y et al (2020) Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368:779–782. <https://doi.org/10.1126/science.abb7498>
- Imbert I et al (2006) A second, non-canonical RNA-dependent RNA polymerase in SARS coronavirus. *EMBO J* 25:4933–4942. <https://doi.org/10.1038/sj.emboj.7601368>
- Johnson MA, Jaudzems K, Wuthrich K (2010) NMR structure of the SARS-CoV nonstructural protein 7 in solution at pH 6.5. *J Mol Biol* 402:619–628. <https://doi.org/10.1016/j.jmb.2010.07.043>
- Kirchdoerfer RN, Ward AB (2019) Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat Commun* 10:2342. <https://doi.org/10.1038/s41467-019-10280-3>
- Konkolova E, Klima M, Nencka R, Boura E (2020) Structural analysis of the putative SARS-CoV-2 primase complex. *J Struct Biol* 211:107548. <https://doi.org/10.1016/j.jsb.2020.107548>
- Lee W, Bahrami A, Dashti HT, Eghbalnia HR, Tonelli M, Westler WM, Markley JL (2019) I-PINE web server: an integrative probabilistic NMR assignment system for proteins. *J Biomol NMR* 73:213–222. <https://doi.org/10.1007/s10858-019-00255-3>
- Lee W, Markley JL (2018) PINE-SPARKY.2 for automated NMR-based protein structure research. *Bioinformatics* 34:1586–1588. <https://doi.org/10.1093/bioinformatics/btx785>
- Lee W, Tonelli M, Markley JL (2015) NMRFAM-SPARKY: enhanced software for biomolecular NMR spectroscopy. *Bioinformatics* 31:1325–1327. <https://doi.org/10.1093/bioinformatics/btu830>
- Peti W et al (2005) Structural genomics of the severe acute respiratory syndrome coronavirus: nuclear magnetic resonance structure of the protein nsP7. *J Virol* 79:12905–12913. <https://doi.org/10.1128/JVI.79.20.12905-12913.2005>
- Shen Y, Bax A (2013) Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *J Biomol NMR* 56:227–241. <https://doi.org/10.1007/s10858-013-9741-y>
- Snijder EJ, Decroly E, Ziebuhr J (2016) The nonstructural proteins directing coronavirus RNA synthesis and processing. *Adv Virus Res* 96:59–126. <https://doi.org/10.1016/bs.aivir.2016.08.008>
- Subissi L et al (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. <https://doi.org/10.1073/pnas.1323705111>
- te Velthuis AJ, van den Worm SH, Snijder EJ (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. <https://doi.org/10.1093/nar/gkr893>
- van der Hoek L (2007) Human coronaviruses: what do they cause? *Antivir Ther* 12:651–658
- Wishart DS et al (1995) ^1H , ^{13}C and ^{15}N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140
- Wu F et al (2020) A new coronavirus associated with human respiratory disease in China. *Nature* 579:265–269. <https://doi.org/10.1038/s41586-020-2008-3>
- Xiao Y et al (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. <https://doi.org/10.1128/JVI.06635-11>
- Ying J, Delaglio F, Torchia DA, Bax A (2017) Sparse multidimensional iterative lineshape-enhanced (SMILE) reconstruction of both non-uniformly sampled and conventional NMR data. *J Biomol NMR* 68:101–118. <https://doi.org/10.1007/s10858-016-0072-7>
- 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. <https://doi.org/10.1038/nsmb999>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

OPEN ACCESS



Edited by:

Qian Han,
Hainan University, China

Reviewed by:

David Douglas Boehr,
Pennsylvania State University, United States
Luis G. Bribea,
National Polytechnic Institute of Mexico
(CINVESTAV), Mexico

***Correspondence:**

Anja Böckmann
a.boeckmann@ibcp.fr
Harald Schwalbe
schwalbe@nmr.uni-frankfurt.de
Martin Hengesbach
hengesch@nmr.uni-frankfurt.de
Andreas Schlundt
schlundt@bio.uni-frankfurt.de

[†]These authors have contributed equally to this work
and share first authorship

[‡]These authors share last authorship

Specialty section:

This article was submitted to
Structural Biology,
a section of the journal
Frontiers in Molecular Biosciences

Received: 13 January 2021

Accepted: 04 February 2021

Published: 10 May 2021

Citation:

Altıncak N, Kom SM, Qureshi NS, Dujardin M, Ninot-Pedrosa M, Abele R, Abi Saad MJ, Alfano C, Almeida FCL, Alshamleh I, de Amorim GC, Anderson TK, Anobom CD, Anorma C, Bains JK, Bax A, Blackledge M, Blechar J, Böckmann A, Brigandat L, Bula A, Bütikofer M, Camacho-Zarco AP, Carlomagno T, Caruso IP, Ceylan B, Chaikuad A, Chu F, Cole L, Crosby MG, de Jesus V, Dhamotharan K, Felli IC, Ferner J, Fleischmann Y, Fogeron M-L, Fourkotis NK, Fuks C, Fürtig B, Gallo A, Gande SL, Gerez JA, Ghosh D, Gomes-Neto F, Gorbatyuk O, Guseva S, Hacker C, Häfner S, Hao B, Hargittay B, Henzler-Wildman K, Hoch JC, Hohmann KF, Hutchison MT, Jaudzems K, Jović K, Kaderli J, Kalniņš G, Kaņepe I, Kirchoerfer RN, Kirkpatrick J, Knapp S, Krishnathas R, Kutz F, zur Lage S, Lambert R, Lang A, Laurents D, Lecoq L, Linhard V, Löhner F, Malki A, Bessa LM, Martin RW, Matzel T, Maurin D, McNutt SW, Mebus-Antunes NC, Meier BH, Meiser N, Mompeán M, Monaca E, Montserret R, Mariño Perez L, Moser C, Muhle-Goll C, Neves-Martins TC, Ni X, Norton-Baker B, Pierattelli R, Pontoriero L, Pustovalova Y, Ohlenschläger O, Orts J, Da Poian AT, Pyper DJ, Richter C, Riek R, Riesenra CM, Robertson A, Pinheiro AS, Sabbatella R, Salvi N, Saxena K, Schulte L, Schiavina M, Schwalbe H, Silber M, Almeida MdS, Sprague-Piercy MA, Spyroulas GA, Sreeramulu S, Tants J-N, Tars K, Torres F, Töws S, Treviño MA, Trucks S, Tsika AC, Varga K, Wang Y, Weber ME, Weigand JE, Wiedemann C, Wimer-Bartoschek J, Wirtz Martin MA, Zehnder J, Hengesbach M and Schlundt A (2021) Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications. *Front. Mol. Biosci.* 8:653148. doi: 10.3389/fmolb.2021.653148

Large-Scale Recombinant Production of the SARS-CoV-2 Proteome for High-Throughput and Structural Biology Applications

Nadide Altıncak^{1,2†}, Sophie Marianne Korn^{2,3†}, Nusrat Shahin Qureshi^{1,2†}, Marie Dujardin^{4†}, Marti Ninot-Pedrosa^{4†}, Rupert Abele⁵, Marie Jose Abi Saad⁶, Caterina Alfano⁷, Fabio C. L. Almeida^{8,9}, Islam Alshamleh^{1,2}, Gisele Cardoso de Amorim^{8,10}, Thomas K. Anderson¹¹, Cristiane D. Anobom^{8,12}, Chelsea Anorma¹³, Jasleen Kaur Bains^{1,2}, Adriaan Bax¹⁴, Martin Blackledge¹⁵, Julius Blechar^{1,2}, Anja Böckmann^{4,*†}, Louis Brigandat⁴, Anna Bula¹⁶, Matthias Bütikofer⁶, Aldo R. Camacho-Zarco¹⁵, Teresa Carlomagno^{17,18}, Icaro Putinhon Caruso^{8,9,19}, Betül Ceylan^{1,2}, Apirat Chaikuad^{20,21}, Feixia Chu²², Laura Cole⁴, Marquise G. Crosby²³, Vanessa de Jesus^{1,2}, Karthikeyan Dhamotharan^{2,3}, Isabella C. Felli^{24,25}, Jan Ferner^{1,2}, Yanick Fleischmann⁶, Marie-Laure Fogeron⁴, Nikolaos K. Fourkotis²⁶, Christin Fuks¹, Boris Fürtig^{1,2}, Angelo Gallo²⁶, Santosh L. Gande^{1,2}, Juan Atilio Gerez⁶, Dhiman Ghosh⁶, Francisco Gomes-Neto^{8,27}, Oksana Gorbatyuk²⁸, Serafima Guseva¹⁵, Carolin Hacker²⁹, Sabine Häfner³⁰, Bing Hao²⁸, Bruno Hargittay^{1,2}, K. Henzler-Wildman¹¹, Jeffrey C. Hoch²⁸, Katharina F. Hohmann^{1,2}, Marie T. Hutchison^{1,2}, Kristaps Jaudzems¹⁶, Katarina Jović²², Janina Kaderli⁶, Gints Kalniņš³¹, Iveta Kaņepe¹⁶, Robert N. Kirchoerfer¹¹, John Kirkpatrick^{17,18}, Stefan Knapp^{20,21}, Robin Krishnathas^{1,2}, Felicitas Kutz^{1,2}, Susanne zur Lage¹⁸, Roderick Lambert³, Andras Lang³⁰, Douglas Laurents³², Lauriane Lecoq⁴, Verena Linhard^{1,2}, Frank Löhner^{2,33}, Anas Malki¹⁵, Luiza Mamigonian Bessa¹⁵, Rachel W. Martin^{13,23}, Tobias Matzel^{1,2}, Damien Maurin¹⁵, Seth W. McNutt²², Nathane Cunha Mebus-Antunes^{8,9}, Beat H. Meier⁶, Nathalie Meiser¹, Miguel Mompeán³², Elisa Monaca⁷, Roland Montserret⁴, Laura Mariño Perez¹⁵, Celine Moser³⁴, Claudia Muhle-Goll³⁴, Thais Cristina Neves-Martins^{8,9}, Xiamonin Ni^{20,21}, Brenna Norton-Baker¹³, Roberta Pierattelli^{24,25}, Letizia Pontoriero^{24,25}, Yulia Pustovalova²⁸, Oliver Ohlenschläger³⁰, Julien Orts⁶, Andrea T. Da Poian⁹, Dennis J. Pyper^{1,2}, Christian Richter^{1,2}, Roland Riek⁶, Chad M. Riesenra³⁵, Angus Robertson¹⁴, Anderson S. Pinheiro^{8,12}, Raffaele Sabbatella⁷, Nicola Salvi¹⁵, Krishna Saxena^{1,2}, Linda Schulte^{1,2}, Marco Schiavina^{24,25}, Harald Schwalbe^{1,2,*†}, Mara Silber³⁴, Marcus da Silva Almeida^{8,9}, Marc A. Sprague-Piercy²³, Georgios A. Spyroulas²⁶, Sridhar Sreeramulu^{1,2}, Jan-Niklas Tants^{2,3}, Kaspars Tars³¹, Felix Torres⁶, Sabrina Töws³, Miguel Á. Treviño³², Sven Trucks¹, Aikaterini C. Tsika²⁶, Krisztina Varga²², Ying Wang¹⁷, Marco E. Weber⁶, Julia E. Weigand³⁶, Christoph Wiedemann³⁷, Julia Wimer-Bartoschek^{1,2}, Maria Alexandra Wirtz Martin^{1,2}, Johannes Zehnder⁶, Martin Hengesbach^{1,*†} and Andreas Schlundt^{2,3,*†}

¹Institute for Organic Chemistry and Chemical Biology, Goethe University Frankfurt, Frankfurt am Main, Germany, ²Center of Biomolecular Magnetic Resonance (BMRZ), Goethe University Frankfurt, Frankfurt am Main, Germany, ³Institute for Molecular Biosciences, Goethe University Frankfurt, Frankfurt am Main, Germany, ⁴Molecular Microbiology and Structural Biochemistry, UMR 5086, CNRS/Lyon University, Lyon, France, ⁵Institute for Biochemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, ⁶Swiss Federal Institute of Technology, Laboratory of Physical Chemistry, ETH Zurich, Zurich, Switzerland, ⁷Structural Biology and Biophysics Unit, Fondazione Ri.MED, Palermo, Italy, ⁸National Center of Nuclear Magnetic Resonance (CNRMN, CENABIO), Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ⁹Institute of Medical Biochemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ¹⁰Multidisciplinary Center for Research in Biology (NUMPEX), Campus Duque de Caxias Federal University of Rio de Janeiro, Duque de Caxias, Brazil, ¹¹Institute for Molecular Virology, University of Wisconsin-Madison, Madison, WI, United States, ¹²Institute of Chemistry, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil, ¹³Department of Chemistry, University of California, Irvine, CA, United States, ¹⁴LCP, NIDDK, NIH, Bethesda, MD, United States, ¹⁵Univ. Grenoble Alpes, CNRS, CEA, IBS, Grenoble, France, ¹⁶Latvian Institute of Organic Synthesis, Riga, Latvia, ¹⁷BMWZ and Institute of Organic

Chemistry, Leibniz University Hannover, Hannover, Germany, ¹⁸Group of NMR-Based Structural Chemistry, Helmholtz Centre for Infection Research, Braunschweig, Germany, ¹⁹Multiuser Center for Biomolecular Innovation (CMIB), Department of Physics, São Paulo State University (UNESP), São José do Rio Preto, Brazil, ²⁰Institute of Pharmaceutical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, ²¹Structural Genomics Consortium, Buchmann Institute for Molecular Life Sciences, Frankfurt am Main, Germany, ²²Department of Molecular, Cellular, and Biomedical Sciences, University of New Hampshire, Durham, NH, United States, ²³Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, United States, ²⁴Magnetic Resonance Centre (CERM), University of Florence, Sesto Fiorentino, Italy, ²⁵Department of Chemistry "Ugo Schiff", University of Florence, Sesto Fiorentino, Italy, ²⁶Department of Pharmacy, University of Patras, Patras, Greece, ²⁷Laboratory of Toxinology, Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil, ²⁸Department of Molecular Biology and Biophysics, UConn Health, Farmington, CT, United States, ²⁹Signals GmbH & Co. KG, Frankfurt am Main, Germany, ³⁰Leibniz Institute on Aging—Fritz Lipmann Institute (FLI), Jena, Germany, ³¹Latvian Biomedical Research and Study Centre, Riga, Latvia, ³²"Rocasolano" Institute for Physical Chemistry (IQFR), Spanish National Research Council (CSIC), Madrid, Spain, ³³Institute of Biophysical Chemistry, Goethe University Frankfurt, Frankfurt am Main, Germany, ³⁴IBG-4, Karlsruhe Institute of Technology, Karlsruhe, Germany, ³⁵Department of Biochemistry and National Magnetic Resonance Facility at Madison, University of Wisconsin-Madison, Madison, WI, United States, ³⁶Department of Biology, Technical University of Darmstadt, Darmstadt, Germany, ³⁷Institute of Biochemistry and Biotechnology, Charles Tanford Protein Centre, Martin Luther University Halle-Wittenberg, Halle/Saale, Germany

The highly infectious disease COVID-19 caused by the *Betacoronavirus* SARS-CoV-2 poses a severe threat to humanity and demands the redirection of scientific efforts and criteria to organized research projects. The international COVID19-NMR consortium seeks to provide such new approaches by gathering scientific expertise worldwide. In particular, making available viral proteins and RNAs will pave the way to understanding the SARS-CoV-2 molecular components in detail. The research in COVID19-NMR and the resources provided through the consortium are fully disclosed to accelerate access and exploitation. NMR investigations of the viral molecular components are designated to provide the essential basis for further work, including macromolecular interaction studies and high-throughput drug screening. Here, we present the extensive catalog of a holistic SARS-CoV-2 protein preparation approach based on the consortium's collective efforts. We provide protocols for the large-scale production of more than 80% of all SARS-CoV-2 proteins or essential parts of them. Several of the proteins were produced in more than one laboratory, demonstrating the high interoperability between NMR groups worldwide. For the majority of proteins, we can produce isotope-labeled samples of HSQC-grade. Together with several NMR chemical shift assignments made publicly available on covid19-nmr.com, we here provide highly valuable resources for the production of SARS-CoV-2 proteins in isotope-labeled form.

Keywords: COVID-19, SARS-CoV-2, nonstructural proteins, structural proteins, accessory proteins, intrinsically disordered region, cell-free protein synthesis, NMR spectroscopy

INTRODUCTION

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2, SCoV2) is the cause of the early 2020 pandemic coronavirus lung disease 2019 (COVID-19) and belongs to *Betacoronaviruses*, a genus of the Coronaviridae family covering the α - δ genera (Leao et al., 2020). The large RNA genome of SCoV2 has an intricate, highly condensed arrangement of coding sequences (Wu et al., 2020). Sequences starting with the main start codon contain an open reading frame 1 (ORF1), which codes for two distinct, large polypeptides (pp), whose relative abundance is governed by the action of an RNA pseudoknot structure element. Upon RNA folding, this element causes a -1 frameshift to allow the continuation of translation, resulting in the generation of a 7,096-amino acid 794 kDa polypeptide. If the pseudoknot is not formed, expression of the first ORF generates a 4,405-amino acid 490 kDa polypeptide. Both the short and long polypeptides translated from this ORF (pp1a and pp1ab, respectively) are posttranslationally cleaved by virus-encoded

proteases into functional, nonstructural proteins (nsps). ORF1a encodes eleven nsps, and ORF1ab additionally encodes the nsps 12–16. The downstream ORFs encode structural proteins (S, E, M, and N) that are essential components for the synthesis of new virus particles. In between those, additional proteins (accessory/auxiliary factors) are encoded, for which sequences partially overlap (Finkel et al., 2020) and whose identification and classification are a matter of ongoing research (Nelson et al., 2020; Pavesi, 2020). In total, the number of identified peptides or proteins generated from the viral genome is at least 28 on the evidence level, with an additional set of smaller proteins or peptides being predicted with high likelihood.

High-resolution studies of SCoV and SCoV2 proteins have been conducted using all canonical structural biology approaches, such as X-ray crystallography on proteases (Zhang et al., 2020) and methyltransferases (MTase) (Krafcikova et al., 2020), cryo-EM of the RNA polymerase (Gao et al., 2020; Yin et al., 2020), and liquid-state (Almeida et al., 2007; Serrano et al., 2009; Cantini et al., 2020; Gallo et al., 2020; Korn et al., 2020a; Korn et al., 2020b;

TABLE 1 | SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

Protein genome position (nt) ^a	Trivial name construct expressed	Size (aa)	Boundaries	MW (kDa)	Homol. SCoV (%) ^b	Template PDB ^c	SCoV2 PDB ^d
nsp1 266–805	Leader	180		19.8	84		
	Full-length	180	1–180	19.8	83		
	Globular domain (GD)	116	13–127	12.7	85	2GDT	7K7P
nsp2 806–2,719		638		70.5	68		
	C-terminal IDR (CtDR)	45	557–601	4.9	55		
nsp3 2,720–8,554		1,945		217.3	76		
a	Ub-like (Ubl) domain	111	1–111	12.4	79	2IDY	7KAG
a	Ub-like (Ubl) domain + IDR	206	1–206	23.2	58		
b	Macrodomain	170	207–376	18.3	74	6VXS	6VXS
c	SUD-N	140	409–548	15.5	69	2W2G	
c	SUD-NM	267	409–675	29.6	74	2W2G	
c	SUD-M	125	551–675	14.2	82	2W2G	
c	SUD-MC	195	551–743	21.9	79	2KQV	
c	SUD-C	64	680–743	7.4	73	2KAF	
d	Papain-like protease PL ^{pro}	318	743–1,060	36	83	6W9C	6W9C
e	NAB	116	1,088–1,203	13.4	87	2K87	
Y	CoV-Y	308	1,638–1,945	34	89		
nsp5 10,055–10,972	Main protease (M^{pro})	306		33.7	96		
	Full-length ^e	306	1–306	33.7	96	6Y84	6Y84
nsp7 11,843–12,091		83		9.2	99		
	Full-length	83	1–83	9.2	99	6WIQ	6WIQ
nsp8 12,092–12,685		198		21.9	98		
	Full-length	198	1–198	21.9	97	6WIQ	6WIQ
nsp9 12,686–13,024		113		12.4	97		
	Full-length	113	1–113	12.4	97	6W4B	6W4B
nsp10 13,025–13,441		139		14.8	97		
	Full-length	139	1–139	14.8	97	6W4H	6W4H
nsp13 16,237–18,039	Helicase	601		66.9	100		
	Full-length	601	1–601	66.9	100	6ZSL	6ZSL
nsp14 18,040–19,620	Exonuclease/ methyltransferase	527		59.8	95		
	Full-length	527	1–527	59.8	95	5NFY	
	MTase domain	240	288–527	27.5	95		
nsp15 19,621–20,658	Endonuclease	346		38.8	89		
	Full-length	346	1–346	38.8	89	6W01	6W01
nsp16 20,659–21,552	Methyltransferase	298		33.3	93		
	Full-length	298	1–298	33.3	93	6W4H	6W4H
ORF3a 25,393–26,220		275		31.3	72		
	Full-length	275	1–275	31.3	72	6XDC	6XDC
ORF4 26,245–26,472	Envelope (E) protein	75		8.4	95		
	Full-length	75	1–75	8.4	95	5X29	7K3G
ORF5 26,523–27,387	Membrane glycoprotein (M)	222		25.1	91		
	Full-length	222	1–222	25.1	91		
ORF6 27,202–27,387		61		7.3	69		
	Full-length	61	1–61	7.3	69		

(Continued on following page)

TABLE 1 | (Continued) SCoV2 protein constructs expressed and purified, given with the genomic position and corresponding PDBs for construct design.

Protein genome position (nt) ^a	Trivial name construct expressed	Size (aa)	Boundaries	MW (kDa)	Homol. SCoV (%) ^b	Template PDB ^c	SCoV2 PDB ^d
ORF7a 27,394–27,759		121		13.7	85		
	Ectodomain (ED)	66	16–81	7.4	85	1XAK	6W37
ORF7b 27,756–27,887		43		5.2	85		
	Full-length	43	1–43	5.2	85		
ORF8 27,894–28,259		121		13.8	32		
	Full-length	121	1–121	13.8	32		
	w/o signal peptide	106	16–121	12	41	7JTL	7JTL
ORF9a 28,274–29,533	Nucleocapsid (N)	419		45.6	91		
	IDR1-NTD-IDR2	248	1–248	26.5	90		
	NTD-SR	169	44–212	18.1	92		
	NTD	136	44–180	14.9	93	6YI3	6YI3
	CTD	118	247–364	13.3	96	2JW8	7C22
ORF9b 28,284–28,574		97		10.8	72		
	Full-length	97	1–97	10.8	72	6Z4U	6Z4U
ORF14 28,734–28,952		73		8	n.a		
	Full-length	73	1–73	8	n.a		
ORF10 29,558–29,674		38		4.4	29		
	Full-length	38	1–38	4.4	29		

^aGenome position in nt corresponding to SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3.

^bSequence identities to SCoV are calculated from an alignment with corresponding protein sequences based on the genome sequence of NCBI Reference NC_004718.3.

^cRepresentative PDB that was available at the beginning of construct design, either SCoV or SCoV2.

^dRepresentative PDB available for SCoV2 (as of December 2020).

^eAdditional point mutations in fl-construct have been expressed.

n.a.: not applicable.

Kubatova et al., 2020; Tonelli et al., 2020) and solid-state NMR spectroscopy of transmembrane (TM) proteins (Mandala et al., 2020). These studies have significantly improved our understanding on the functions of molecular components, and they all rely on the recombinant production of viral proteins in high amount and purity.

Apart from structures, purified SCoV2 proteins are required for experimental and preclinical approaches designed to understand the basic principles of the viral life cycle and processes underlying viral infection and transmission. Approaches range from studies on immune responses (Esposito et al., 2020), antibody identification (Jiang et al., 2020), and interactions with other proteins or components of the host cell (Bojkova et al., 2020; Gordon et al., 2020). These examples highlight the importance of broad approaches for the recombinant production of viral proteins.

The research consortium *COVID19-NMR* founded in 2020 seeks to support the search for antiviral drugs using an NMR-based screening approach. This requires the large-scale production of all druggable proteins and RNAs and their NMR resonance assignments. The latter will enable solution structure determination of viral proteins and RNAs for rational drug design and the fast mapping of compound binding sites. We have recently produced and determined secondary structures of SCoV2 RNA *cis*-regulatory elements in near completeness by NMR spectroscopy, validated by DMS-

MaPseq (Wacker et al., 2020), to provide a basis for RNA-oriented fragment screens with NMR.

We here compile a compendium of more than 50 protocols (see **Supplementary Tables S11–S123**) for the production and purification of 23 of the 30 SCoV2 proteins or fragments thereof (summarized in **Tables 1, 2**). We defined those 30 proteins as existing or putative ones to our current knowledge (see later discussion). This compendium has been generated in a coordinated and concerted effort between >30 labs worldwide (**Supplementary Table S1**), with the aim of providing pure mg amounts of SCoV2 proteins. Our protocols include the rational strategy for construct design (if applicable, guided by available homolog structures), optimization of expression, solubility, yield, purity, and suitability for follow-up work, with a focus on uniform stable isotope-labeling.

We also present protocols for a number of accessory and structural E and M proteins that could only be produced using wheat-germ cell-free protein synthesis (WG-CFPS). In SCoV2, accessory proteins represent a class of mostly small and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems. They are often found in inclusion bodies and difficult to purify in quantities adequate for structural studies. We thus here exploit cell-free synthesis, mainly based on previous reports on production and purification of viral membrane proteins in general (Fogeron et al., 2015b; Fogeron et al., 2017; Jirasko

TABLE 2 | Summary of SCoV2 protein production results in *Covid19-NMR*.

Construct expressed	Yields (mg/L) ^a or (mg/ml) ^b	Results	Comments	BMRB	Supplementary Material
nsp1					
fl	5	NMR assigned	Expression only at >20°C; after 7 days at 25°C partial proteolysis	50620 ^d	SI1
GD	>0.5	HSQC	High expression; mainly insoluble; higher salt increases stability (>250 mM)		
nsp2					
CtDR	0.7–1.5	NMR assigned	Assignment with His-tag shown in (Mompéan et al., 2020)	50687 ^c	SI2
nsp3					
UBI	0.7	HSQC	Highly stable over weeks; spectrum overlays with Ubl + IDR		SI3
UBI + IDR	2–3	NMR assigned	Highly stable for >2 weeks at 25°C	50446 ^d	
Macrodomain	9	NMR assigned	Highly stable for >1 week at 25°C and > 2 weeks at 4°C	50387 ^d 50388 ^d	
SUD-N	14	NMR assigned	Highly stable for >10 days at 25°C	50448 ^d	
SUD-NM	17	HSQC	Stable for >1 week at 25°C		
SUD-M	8.5	NMR assigned	Significant precipitation during measurement; tendency to dimerize	50516 ^d	
SUD-MC	12	HSQC	Stable for >1 week at 25°C		
SUD-C	4.7	NMR assigned	Stable for >10 days at 25°C	50517 ^d	
PL ^{pro}	12	HSQC	Solubility-tag essential for expression; tendency to aggregate		
NAB	3.5	NMR assigned	Highly stable for >1 week at 25°C; stable for >5 weeks at 4°C	50334 ^d	
CoV-Y	12	HSQC	Low temperature (<25°C) and low concentrations (<0.2 mM) favor stability; gradual degradation at 25°C; lithium bromide in final buffer supports solubility		
nsp5					
fl	55	HSQC	Impaired dimerization induced by artificial N-terminal residues		SI4
nsp7					
fl	17	NMR assigned	Stable for several days at 35°C; stable for >1 month at 4°C	50337 ^d	SI5
nsp8					
fl	17	HSQC	Concentration dependent aggregation; low concentrations favor stability		SI6
nsp9					
fl	4.5	NMR assigned	Stable dimer for >4 months at 4°C and >2 weeks at 25°C	50621 ^d 50622 ^d 50513	SI7
nsp10					
fl	15	NMR assigned	Zn ²⁺ addition during expression and purification increases protein stability; stable for >1 week at 25°C	50392	SI8
nsp13					
fl	0.5	HSQC	Low expression; protein unstable; concentration above 20 µM not possible		SI9
nsp14					
fl	6	Pure protein	Not above 50 µM; best storage: with 50% (v/v) glycerol; addition of reducing agents		SI10
MTase	10	Pure protein	As fl nsp14; high salt (>0.4 M) for increased stability; addition of reducing agents		
nsp15					
fl	5	HSQC	Tendency to aggregate at 25°C		SI11
nsp16					
fl	10	Pure protein	Addition of reducing agents; 5% (v/v) glycerol favorable; highly unstable		SI12
ORF3a					
fl	0.6	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		SI13
E protein					
fl	0.45	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		SI14

(Continued on following page)

TABLE 2 | (Continued) Summary of SCoV2 protein production results in *Covid19-NMR*.

Construct expressed	Yields (mg/L) ^a or (mg/ml) ^b	Results	Comments	BMRB	Supplementary Material
M Protein					SI15
fl	0.33	Pure protein	Addition of detergent during expression (0.05% Brij-58); stable protein		
ORF6					SI16
fl	0.27	HSQC	Soluble expression without detergent; stable protein; no expression with STREP-tag at N-terminus		
ORF7a					SI17
ED	0.4	HSQC	Unpurified protein tends to precipitate during refolding, purified protein stable for 4 days at 25°C		
ORF7b					SI18
fl	0.6	HSQC	Tendency to oligomerize; solubilizing agents needed		
fl	0.27	HSQC	Addition of detergent during expression (0.1% MNG-3); stable protein		
ORF8					SI19
fl	0.62	HSQC	Tendency to oligomerize		
ΔORF8	0.5	Pure protein			
N protein					SI20
IDR1-NTD- IDR2	12	NMR assigned	High salt (>0.4 M) for increased stability	50618, 50619, 50558, 50557 ^d	
NTD-SR	3	HSQC			
NTD	3	HSQC		34511	
CTD	2	NMR assigned	Stable dimer for >4 months at 4°C and >3 weeks at 30°C	50518 ^d	
ORF9b					SI21
fl	0.64	HSQC	Expression without detergent, protein is stable		
ORF14					SI22
fl	0.43	HSQC	Addition of detergent during expression (0.05% Brij-58); stable in detergent but unstable on lipid reconstitution		
ORF10					SI23
fl	2	HSQC	Tendency to oligomerize; unstable upon tag cleavage		

^aYields from bacterial expression represent the minimal protein amount in mg/L independent of the cultivation medium. *Italic values indicate yields from CFPS.*

^bYields from CFPS represent the minimal protein amount in mg/ml of wheat-germ extract.

^cCOVID19-nmr BMRB depositions yet to be released.

^dCOVID19-nmr BMRB depositions.

et al., 2020b). Besides yields compatible with structural studies, ribosomes in WG extracts further possess an increased folding capacity (Netzer and Hartl, 1997), favorable for those more complicated proteins.

We exemplify in more detail the optimization of protein production, isotope-labeling, and purification for proteins with different individual challenges: the nucleic acid-binding (NAB) domain of nsp3e, the main protease nsp5, and several auxiliary proteins. For the majority of produced and purified proteins, we achieve >95% purity and provide ¹⁵N-HSQC spectra as the ultimate quality measure. We also provide additional suggestions for challenging proteins, where our protocols represent a unique resource and starting point exploitable by other labs.

MATERIALS AND METHODS

Strains, Plasmids, and Cloning

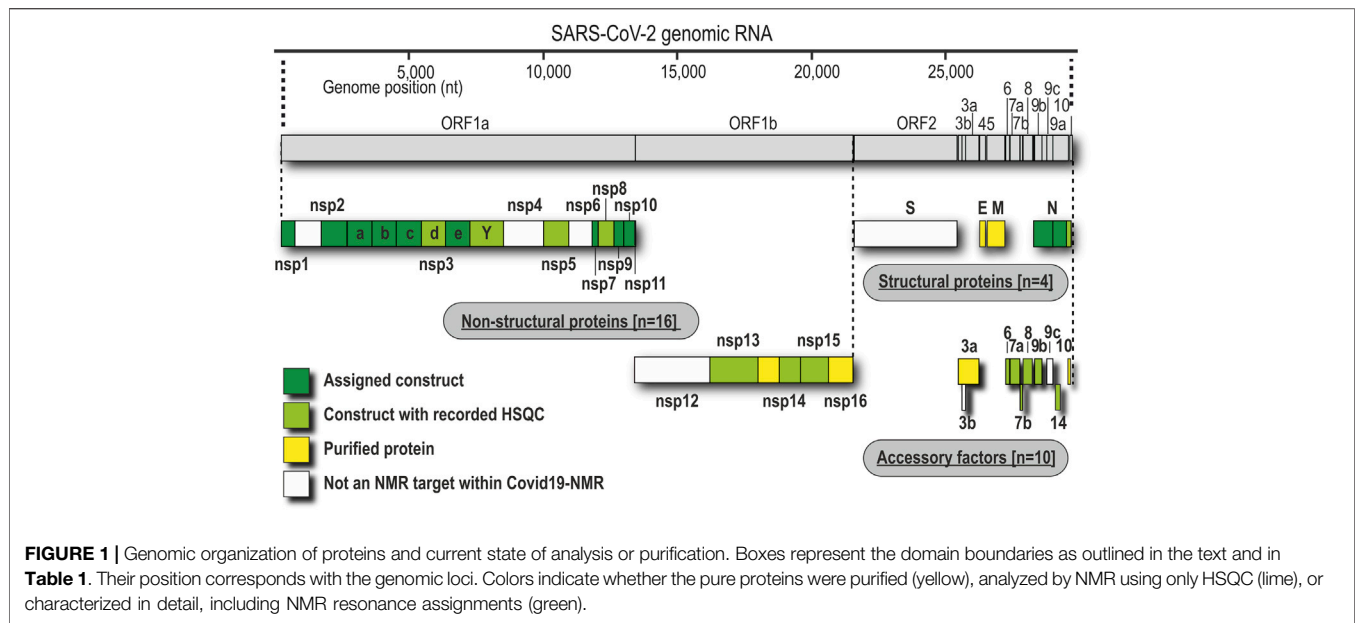
The rationale of construct design for all proteins can be found within the respective protocols in **Supplementary Tables SI1–SI23**. For bacterial production, *E. coli* strains and expression plasmids are given; for WG-CFPS, template

vectors are listed. Protein coding sequences of interest have been obtained as either commercial, codon-optimized genes or, for shorter ORFs and additional sequences, annealed from oligonucleotides prior to insertion into the relevant vector. Subcloning of inserts, adjustment of boundaries, and mutations of genes have been carried out by standard molecular biology techniques. All expression plasmids can be obtained upon request from the *COVID19-NMR* consortium (<https://covid19-nmr.com/>), including information about coding sequences, restriction sites, fusion tags, and vector backbones.

Protein Production and Purification

For SCoV2 proteins, we primarily used heterologous production in *E. coli*. Detailed protocols of individual full-length (fl) proteins, separate domains, combinations, or particular expression constructs as listed in **Table 1** can be found in the (**Supplementary Tables SI1–SI23**).

The ORF3a, ORF6, ORF7b, ORF8, ORF9b, and ORF14 accessory proteins and the structural proteins M and E were produced by WG-CFPS as described in the **Supplementary Material**. In brief, transcription and translation steps have



been performed separately, and detergent has been added for the synthesis of membrane proteins as described previously (Takai et al., 2010; Fogeron et al., 2017).

NMR Spectroscopy

All amide correlation spectra, either HSQC- or TROSY-based, are representative examples. Details on their acquisition parameters and the raw data are freely accessible through <https://covid19-nmr.de> or upon request.

RESULTS

In the following, we provide protocols for the purification of SCoV2 proteins sorted into 1) nonstructural proteins and 2) structural proteins together with accessory ORFs. **Table 1** shows an overview of expression constructs. We use a consequent terminology of those constructs, which is guided by domains, intrinsically disordered regions (IDRs) or other particularly relevant sequence features within them. This study uses the SCoV2 NCBI reference genome entry NC_045512.2, identical to GenBank entry MN908947.3 (Wu et al., 2020), unless denoted differently in the respective protocols. Any relevant definition of boundaries can also be found in the SI protocols.

As applicable for a major part of our proteins, we further define a standard procedure for the purification of soluble His-tagged proteins that are obtained through the sequence of IMAC, TEV/Ulp1 Protease cleavage, Reverse IMAC, and Size-exclusion chromatography, eventually with individual alterations, modifications, or additional steps. For convenient reading, we will thus use the abbreviation IPRS to avoid redundant protocol description. Details for every protein,

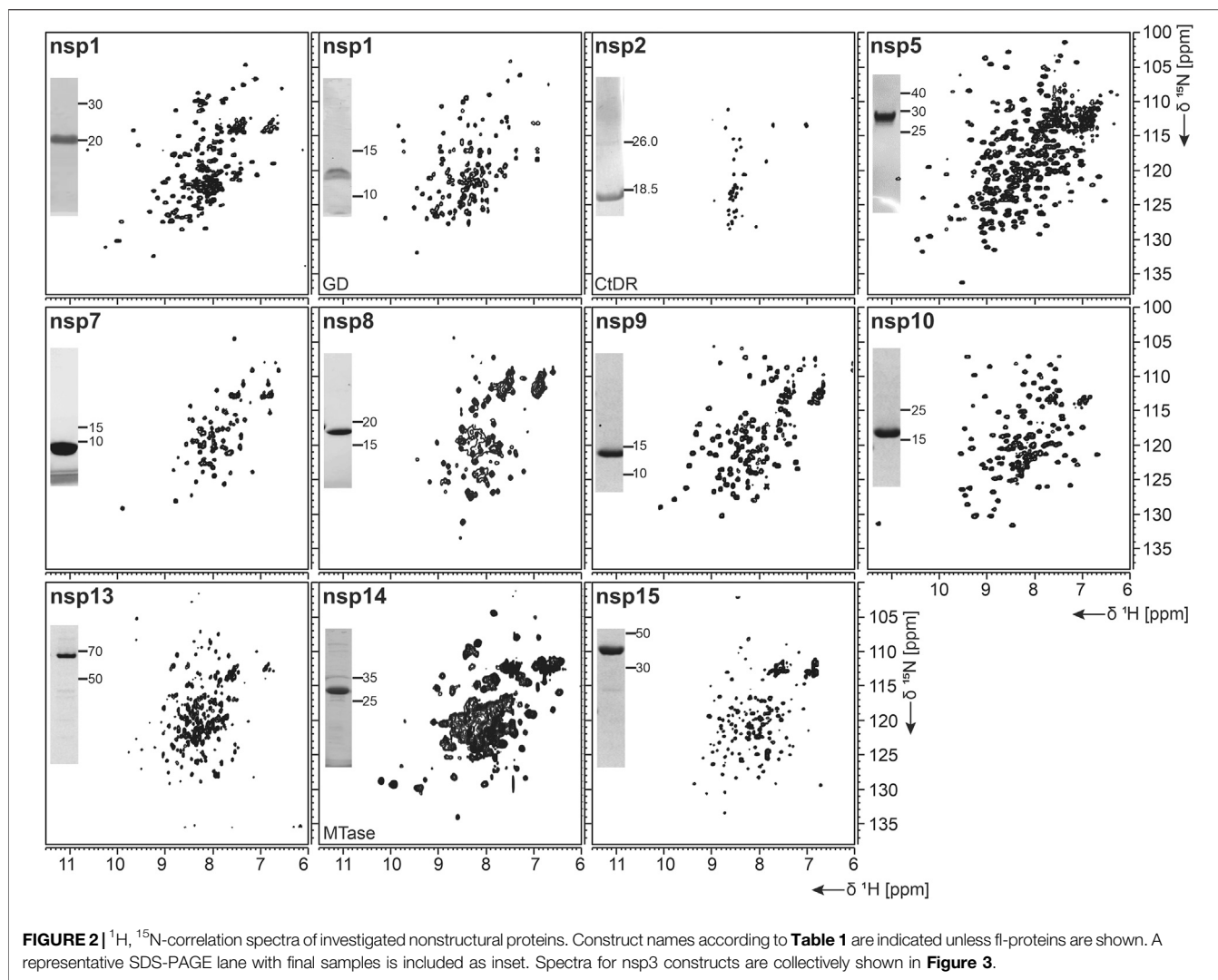
including detailed expression conditions, buffers, incubation times, supplements, storage conditions, yields, and stability, can be found in the respective **Supplementary Tables S11–S123** (see also **Supplementary Tables S1, S2**) and **Tables 1, 2**.

Nonstructural Proteins

We have approached and challenged the recombinant production of a large part of the SCoV2 nsps (**Figure 1**), with great success (**Table 2**). We excluded nsp4 and nsp6 (TM proteins), which are little characterized and do not reveal soluble, folded domains by prediction (Oostra et al., 2007; Oostra et al., 2008). The function of the very short (13 aa) nsp11 is unknown, and it seems to be a mere copy of the nsp12 amino-terminal residues, remaining as a protease cleavage product of ORF1a. Further, we left out the RNA-dependent RNA polymerase nsp12 in our initial approach because of its size (>100 kDa) and known unsuitability for heterologous recombinant production in bacteria. Work on NMR-suitable nsp12 bacterial production is ongoing, while other expert labs have succeeded in purifying nsp12 for cryo-EM applications in different systems (Gao et al., 2020; Hillen et al., 2020). For the remainder of nsps, we here provide protocols for fl-proteins or relevant fragments of them.

nsp1

nsp1 is the very N-terminus of the polyproteins pp1a and pp1ab and one of the most enigmatic viral proteins, expressed only in α - and β -CoVs (Narayanan et al., 2015). Interestingly, nsp1 displays the highest divergence in sequence and size among different CoVs, justifying it as a genus-specific marker (Snijder et al., 2003). It functions as a host shutoff factor by suppressing innate immune functions and host gene expression (Kamitani et al.,



2006; Narayanan et al., 2008; Schubert et al., 2020). This suppression is achieved by an interaction of the nsp1 C-terminus with the mRNA entry tunnel within the 40 S subunit of the ribosome (Schubert et al., 2020; Thoms et al., 2020).

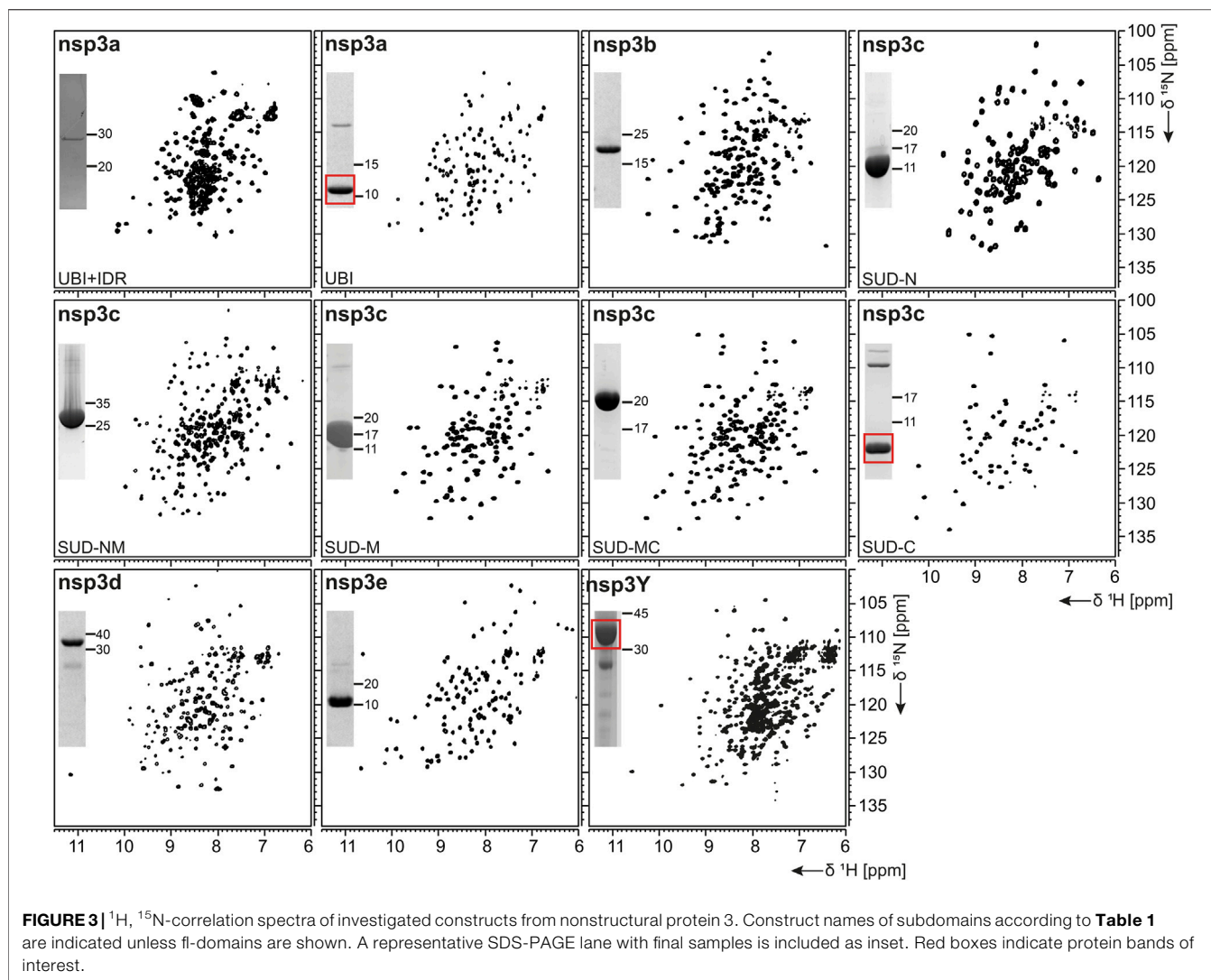
As summarized in **Table 1**, fl-domain boundaries of nsp1 were chosen to contain the first 180 amino acids, in analogy to its closest homolog from SCoV (Snijder et al., 2003). In addition, a shorter construct was designed, encoding only the globular core domain (GD, aa 13–127) suggested by the published SCoV nsp1 NMR structure (Almeida et al., 2007). His-tagged fl nsp1 was purified using the IPRS approach. Protein quality was confirmed by the available HSQC spectrum (**Figure 2**). Despite the flexible C-terminus, we were able to accomplish a near-complete backbone assignment (Wang et al., 2021).

Interestingly, the nsp1 GD was found to be problematic in our hands despite good expression. We observed insolubility, although buffers were used according to the

homolog SCoV nsp1 GD (Almeida et al., 2007). Nevertheless, using a protocol comparable to the one for fl nsp1, we were able to record an HSQC spectrum proving a folded protein (**Figure 2**).

nsp2

nsp2 has been suggested to interact with host factors involved in intracellular signaling (Cornillez-Ty et al., 2009; Davies et al., 2020). The precise function, however, is insufficiently understood. Despite its potential dispensability for viral replication in general, it might be a valuable model to gain insights into virulence due to its possible involvement in the regulation of global RNA synthesis (Graham et al., 2005). We provide here a protocol for the purification of the C-terminal IDR (CtDR) of nsp2 from residues 557 to 601, based on disorder predictions [PrDOS (Ishida and Kinoshita, 2007)]. The His-Trx-tagged peptide was purified by IPRS. Upon dialysis, two IEC steps were performed: first anionic and then cationic, with good final yields (**Table 1**). Stability and



purity were confirmed by an HSQC spectrum (**Figure 2**) and a complete backbone assignment (Mompean et al., 2020; **Table 2**).

nsp3

nsp3, the largest nsp (Snijder et al., 2003), is composed of a plethora of functionally related, yet independent, subunits. After cleavage of nsp3 from the fl ORF1-encoded polypeptide chain, it displays a 1945-residue multidomain protein, with individual functional entities that are subclassified from nsp3a to nsp3e followed by the ectodomain embedded in two TM regions and the very C-terminal CoV-Y domain. The soluble nsp3a-3e domains are linked by various types of linkers with crucial roles in the viral life cycle and are located in the so-called viral cytoplasm, which is separated from the host cell after budding off the endoplasmic reticulum and contains the viral RNA (Wolff et al., 2020). Remarkably, the nsp3c substructure comprises three subdomains, making nsp3

the most complex SCoV2 protein. The precise function and eventual RNA-binding specificities of nsp3 domains are not yet understood. We here focus on the nsp3 domains a–e and provide elaborated protocols for additional constructs carrying relevant linkers or combinations of domains (**Table 1**). Moreover, we additionally present a convenient protocol for the purification of the C-terminal CoV-Y domain.

nsp3a

The N-terminal portion of nsp3 is comprised of a ubiquitin-like (Ubl) structured domain and a subsequent acidic IDR. Besides its ability to bind ssRNA (Serrano et al., 2007), nsp3a has been reported to interact with the nucleocapsid (Hurst et al., 2013; Khan et al., 2020), playing a potential role in virus replication. We here provide protocols for the purification of both the Ubl (aa 1–111) and fl nsp3a (aa 1–206), including the acidic IDR (Ubl + IDR **Table 1**). Domain boundaries were defined similar to the published NMR structure of SCoV nsp3a (Serrano et al., 2007). His-

tagged nsp3a Ubl + IDR and GST-tagged nsp3a Ubl were each purified via the IPRS approach. nsp3a Ubl yielded mM sample concentrations and displayed a well-dispersed HSQC spectrum (**Figure 3**). Notably, the herein described protocol also enables purification of fl nsp3a (Ubl + IDR) (**Tables 1, 2**). Despite the unstructured IDR overhang, the excellent protein quality and stability allowed for near-complete backbone assignment [**Figure 3**, (Salvi et al., 2021)].

nsp3b

nsp3b is an ADP-ribose phosphatase macrodomain and potentially plays a key role in viral replication. Moreover, the de-ADP ribosylation function of nsp3b protects SCoV2 from antiviral host immune response, making nsp3b a promising drug target (Frick et al., 2020). As summarized in **Table 1**, the domain boundaries of the herein investigated nsp3b are residues 207–376 of the nsp3 primary sequence and were identical to available crystal structures with PDB entries 6YWM and 6YWL (unpublished). For purification, we used the IPRS approach, which yielded pure fl nsp3b (**Table 2**). Fl nsp3b displays well-dispersed HSQC spectra, making this protein an amenable target for NMR structural studies. In fact, we recently reported near-to-complete backbone assignments for nsp3b in its apo and ADP-ribose-bound form (Cantini et al., 2020).

nsp3c

The SARS unique domain (SUD) of nsp3c has been described as a distinguishing feature of SCoVs (Snijder et al., 2003). However, similar domains in more distant CoVs, such as MHV or MERS, have been reported recently (Chen et al., 2015; Kusov et al., 2015). nsp3c comprises three distinct globular domains, termed SUD-N, SUD-M, and SUD-C, according to their sequential arrangement: N-terminal (N), middle (M), and C-terminal (C). SUD-N and SUD-M develop a macrodomain fold similar to nsp3b and are described to bind G-quadruplexes (Tan et al., 2009), while SUD-C preferentially binds to purine-containing RNA (Johnson et al., 2010). Domain boundaries for SUD-N and SUD-M and for the tandem-domain SUD-NM were defined in analogy to the SCoV homolog crystal structure (Tan et al., 2009). Those for SUD-C and the tandem SUD-MC were based on NMR solution structures of corresponding SCoV homologs (**Table 1**) (Johnson et al., 2010). SUD-N, SUD-C, and SUD-NM were purified using GST affinity chromatography, whereas SUD-M and SUD-MC were purified using His affinity chromatography. Removal of the tag was achieved by thrombin cleavage and final samples of all domains were prepared subsequent to size-exclusion chromatography (SEC). Except for SUD-M, all constructs were highly stable (**Table 2**). Overall protein quality allowed for the assignment of backbone chemical shifts for the three single domains (Gallo et al., 2020) and good resolved HSQC spectra also for the tandem domains (**Figure 3**).

nsp3d

nsp3d comprises the papain-like protease (PL^{Pro}) domain of nsp3 and, hence, is one of the two SCoV2 proteases that are responsible for processing the viral polypeptide chain and generating functional proteins (Shin et al., 2020). The domain boundaries of PL^{Pro} within nsp3 are set by residues 743 and 1,060 (**Table 1**). The protein is particularly challenging, as it is prone to misfolding and rapid precipitation. We prepared His-tagged and His-SUMO-tagged PL^{Pro}. The His-tagged version mainly remained in the insoluble fraction. Still, mg quantities could be purified from the soluble fraction, however, greatly misfolded. Fusion to SUMO significantly enhanced protein yield of soluble PL^{Pro}. The His-SUMO-tag allowed simple IMAC purification, followed by cleavage with Ulp1 and isolation of cleaved PL^{Pro} via a second IMAC. A final purification step using gel filtration led to pure PL^{Pro} of both unlabeled and 15N-labeled species (**Table 2**). The latter has allowed for the acquisition of a promising amide correlation spectrum (**Figure 3**).

nsp3e

nsp3e is unique to *Betacoronaviruses* and consists of a nucleic acid-binding domain (NAB) and the so-called group 2-specific marker (G2M) (Neuman et al., 2008). Structural information is rare; while the G2M is predicted to be intrinsically disordered (Lei et al., 2018); the only available experimental structure of the nsp3e NAB was solved from SCoV by the Wüthrich lab using solution NMR (Serrano et al., 2009). We here used this structure for a sequence-based alignment to derive reasonable domain boundaries for the SCoV2 nsp3e NAB (**Figures 4A,B**). The high sequence similarity suggested using nsp3 residues 1,088–1,203 (**Table 1**). This polypeptide chain was encoded in expression vectors comprising His- and His-GST tags, both cleavable by TEV protease. Both constructs showed excellent expression, suitable for the IPRS protocol (**Figure 4C**). Finally, a homogenous NAB species, as supported by the final gel of pooled samples (**Figure 4D**), was obtained. The excellent protein quality and stability are supported by the available HSQC (**Figure 3**) and a published backbone assignment (Korn et al., 2020a).

nsp3Y

nsp3Y is the most C-terminal domain of nsp3 and exists in all coronaviruses (Neuman et al., 2008; Neuman, 2016). Together, though, with its preceding regions G2M, TM 1, the ectodomain, TM2, and the Y1-domain, it has evaded structural investigations so far. The precise function of the CoV-Y domain remains unclear, but, together with the Y1-domain, it might affect binding to nsp4 (Hagemeyer et al., 2014). We were able to produce and purify nsp3Y (CoV-Y) comprising amino acids 1,638–1,945 (**Table 1**), yielding 12 mg/L with an optimized protocol that keeps the protein in a final NMR buffer containing HEPES and lithium bromide. Although the protein still shows some tendency to aggregate and degrade (**Table 2**), and despite its relatively large size, the spectral quality is excellent (**Figure 3**). nsp3 CoV-Y appears suitable for an NMR backbone

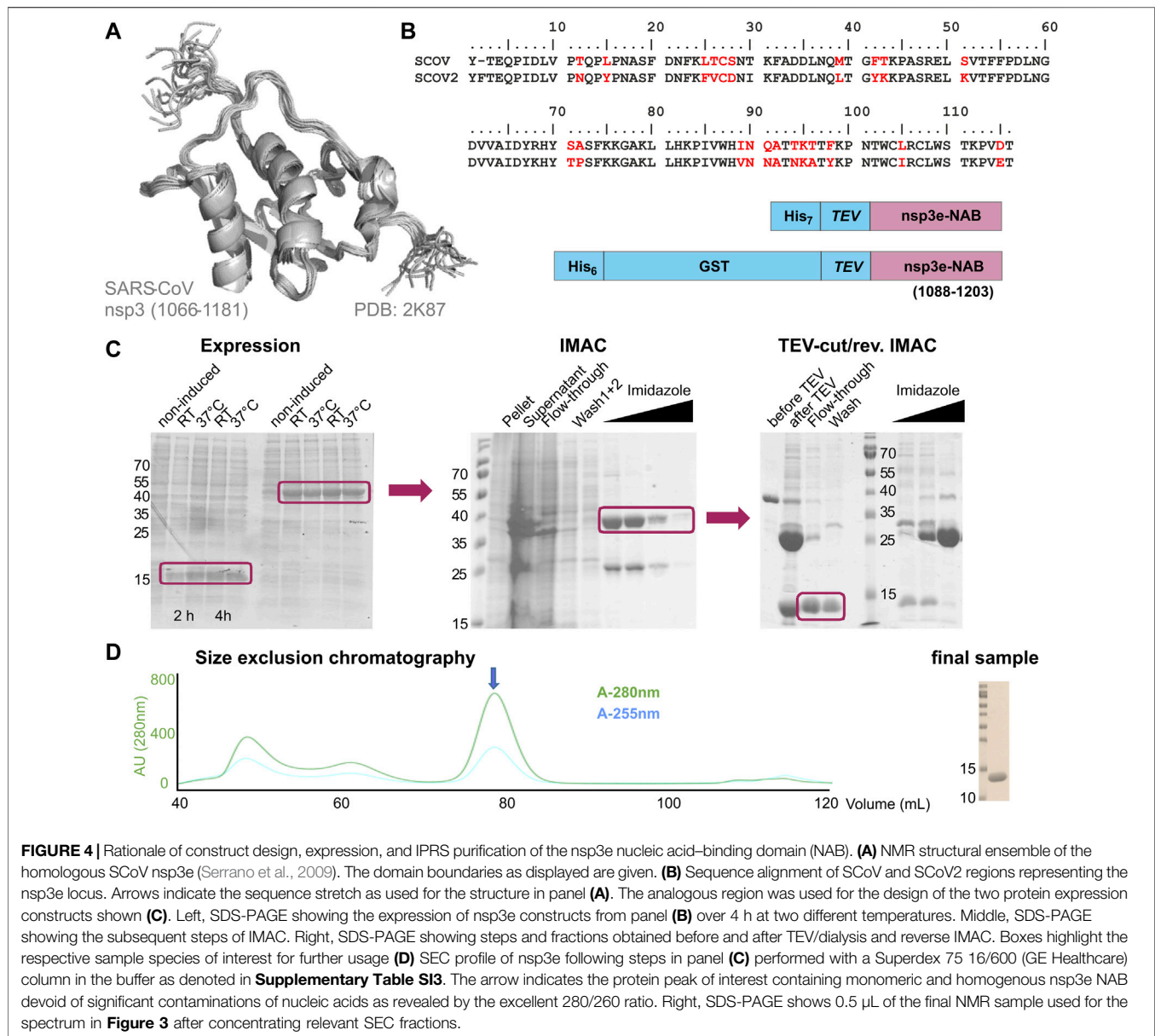


FIGURE 4 | Rationale of construct design, expression, and IPRES purification of the nsp3e nucleic acid-binding domain (NAB). **(A)** NMR structural ensemble of the homologous SCoV nsp3e (Serrano et al., 2009). The domain boundaries as displayed are given. **(B)** Sequence alignment of SCoV and SCoV2 regions representing the nsp3e locus. Arrows indicate the sequence stretch as used for the structure in panel **(A)**. The analogous region was used for the design of the two protein expression constructs shown **(C)**. Left, SDS-PAGE showing the expression of nsp3e constructs from panel **(B)** over 4 h at two different temperatures. Middle, SDS-PAGE showing the subsequent steps of IMAC. Right, SDS-PAGE showing steps and fractions obtained before and after TEV/dialysis and reverse IMAC. Boxes highlight the respective sample species of interest for further usage **(D)** SEC profile of nsp3e following steps in panel **(C)** performed with a Superdex 75 16/600 (GE Healthcare) column in the buffer as denoted in **Supplementary Table S13**. The arrow indicates the protein peak of interest containing monomeric and homogenous nsp3e NAB devoid of significant contaminations of nucleic acids as revealed by the excellent 280/260 ratio. Right, SDS-PAGE shows 0.5 μ L of the final NMR sample used for the spectrum in **Figure 3** after concentrating relevant SEC fractions.

assignment carried out at lower concentrations in a deuterated background (ongoing).

nsp5

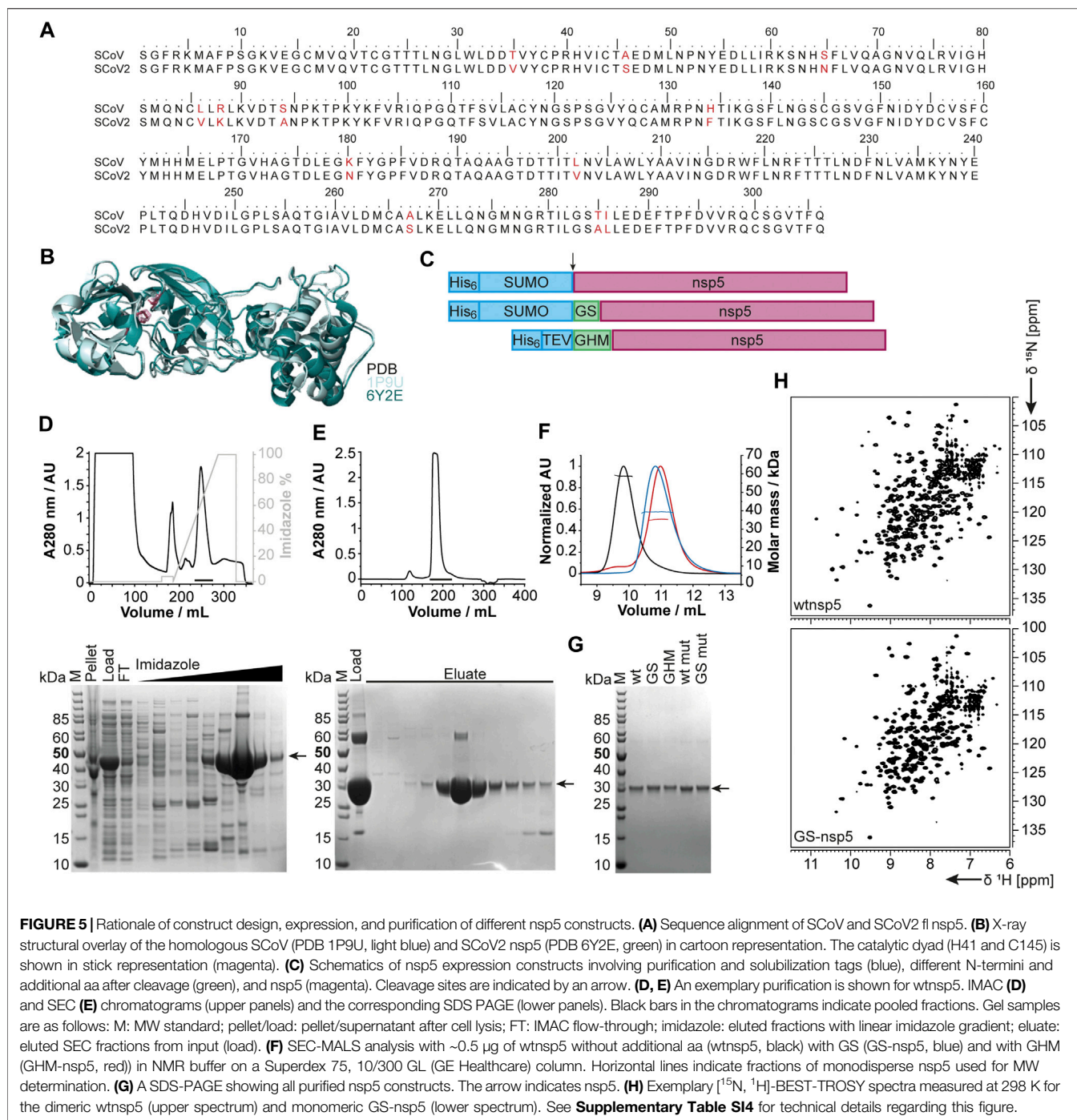
The functional main protease nsp5 (M^{Pro}) is a dimeric cysteine protease (Ullrich and Nitsche, 2020). Amino acid sequence and 3D structure of SCoV [PDB 1P9U (Anand et al., 2003)] and SCoV2 (PDB 6Y2E [Zhang et al., 2020]) homologs are highly conserved (**Figures 5A,B**). The dimer interface involves the N-termini of both monomers, which puts considerable constraints on the choice of protein sequence for construct design regarding the N-terminus.

We thus designed different constructs differing in the N-terminus: the native N-terminus (wt), a GS mutant with the additional N-terminal residues glycine and serine as His-SUMO

fusion, and a GHM mutant with the amino acids glycine, histidine, and methionine located at the N-terminus with His-tag and TEV cleavage site (**Figure 5C**). Purification of all proteins via the IPRES approach (**Figures 5D,E**) yielded homogenous and highly pure protein, analyzed by PAGE (**Figure 5G**), mass spectrometry, and 2D [^{15}N , 1H]-BEST TROSY spectra (**Figure 5H**). Final yields are summarized in **Table 2**.

nsp7 and nsp8

Both nsp7 and nsp8 are auxiliary factors of the polymerase complex together with the RNA-dependent RNA polymerase nsp12 and have high sequence homology with SCoV (100% and 99%, respectively) (Gordon et al., 2020). For nsp7 in complex with nsp8 or for nsp8 alone, additional functions in RNA synthesis priming have been proposed (Tvarogova et al., 2019;



Konkolova et al., 2020). In a recent study including an RNA-substrate-bound structure (Hillen et al., 2020), both proteins (with two molecules of nsp8 and one molecule of nsp7 for each nsp12 RNA polymerase) were found to be essential for polymerase activity in SCoV2. For both fl-proteins, a previously established expression and IPRS purification strategy for the SCoV proteins (Kirchdoerfer and Ward, 2019) was successfully transferred, which resulted in decent yields of

reasonably stable proteins (**Table 2**). Driven by its intrinsically oligomeric state, nsp8 showed some tendency toward aggregation, limiting the available sample concentration. The higher apparent molecular weight and limited solubility are also reflected in the success of NMR experiments. While we succeeded in a complete NMR backbone assignment of nsp7 (Tonelli et al., 2020), the quality of the spectra obtained for nsp8 is currently limited to the HSQC presented in **Figure 2**.

nsp9

The 12.4 kDa ssRNA-binding nsp9 is highly conserved among *Betacoronaviruses*. It is a crucial part of the viral replication machinery (Miknis et al., 2009), possibly targeting the 3'-end stem-loop II (s2m) of the genome (Robertson et al., 2005). nsp9 adopts a fold similar to oligonucleotide/oligosaccharide-binding proteins (Egloff et al., 2004), and structural data consistently uncovered nsp9 to be dimeric in solution (Egloff et al., 2004; Sutton et al., 2004; Miknis et al., 2009; Littler et al., 2020). Dimer formation seems to be a prerequisite for viral replication (Miknis et al., 2009) and influences RNA-binding (Sutton et al., 2004), despite a moderate affinity for RNA *in vitro* (Littler et al., 2020).

Based on the early available crystal structure of SCoV2 nsp9 (PDB 6W4B, unpublished), we used the 113 aa fl sequence of nsp9 for our expression construct (Table 1). Production of either His- or His-GST-tagged fl nsp9 yielded high amounts of soluble protein in both natural abundance and ¹³C- and ¹⁵N-labeled form. Purification *via* the IPRS approach enabled us to separate fl nsp9 in different oligomer states. The earliest eluted fraction represented higher oligomers, was contaminated with nucleic acids and was not possible to concentrate above 2 mg/ml. This was different for the subsequently eluting dimeric fl nsp9 fraction, which had a A260/280 ratio of below 0.7 and could be concentrated to >5 mg/ml (Table 2). The excellent protein quality and stability are supported by the available HSQC (Figure 2), and a near-complete backbone assignment (Dudas et al., 2021).

nsp10

The last functional protein encoded by ORF1a, nsp10, is an auxiliary factor for both the methyltransferase/exonuclease nsp14 and the 2'-O-methyltransferase (MTase) nsp16. However, it is required for the MTase activity of nsp16 (Krafcikova et al., 2020), it confers exonuclease activity to nsp14 in the RNA polymerase complex in SCoV (Ma et al., 2015). It contains two unusual zinc finger motifs (Joseph et al., 2006) and was initially proposed to comprise RNA-binding properties. We generated a construct (Table 1) containing an expression and affinity purification tag on the N-terminus as reported for the SCoV variant (Joseph et al., 2006). Importantly, additional Zn²⁺ ions present during expression and purification stabilize the protein significantly (Kubatova et al., 2020). The yield during isotope-labeling was high (Table 2), and tests in unlabeled rich medium showed the potential for yields exceeding 100 mg/L. These characteristics facilitated in-depth NMR analysis and a backbone assignment (Kubatova et al., 2020).

nsp13

nsp13 is a conserved ATP-dependent helicase that has been characterized as part of the RNA synthesis machinery by binding to nsp12 (Chen et al., 2020b). It represents an interesting drug target, for which the available structure (PDB 6ZSL) serves as an excellent basis (Table 1). The precise molecular function, however, has remained enigmatic since it is not clear whether the RNA unwinding function is required for making ssRNA accessible for RNA synthesis (Jia et al., 2019) or whether it is required for proofreading and backtracking (Chen

et al., 2020b). We obtained pure protein using a standard expression vector, generating a His-SUMO-tagged protein. Following Ulp1 cleavage, the protein showed limited protein stability in the solution (Table 2).

nsp14

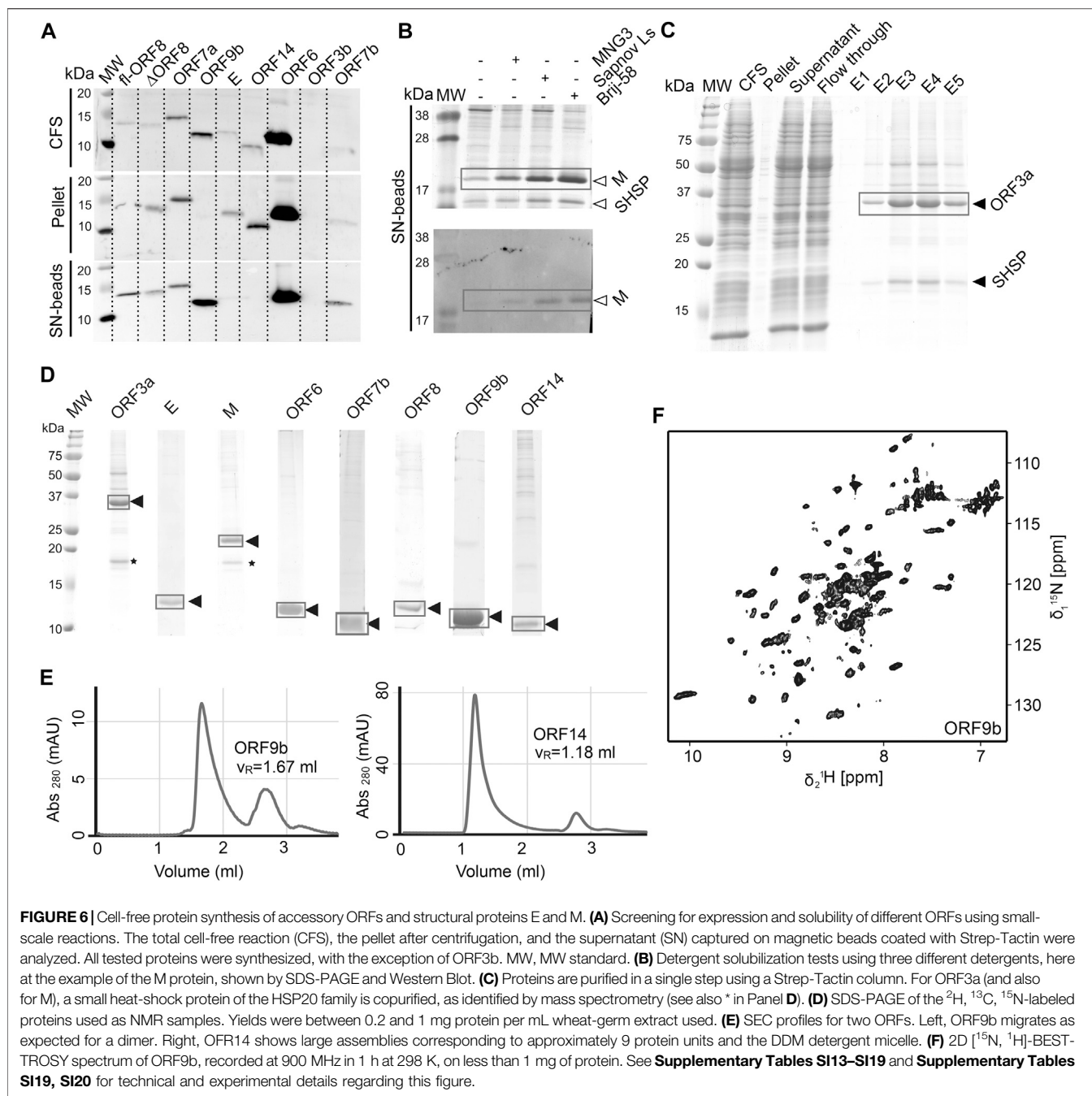
nsp14 contains two domains: an N-terminal exonuclease domain and a C-terminal MTase domain (Ma et al., 2015). The exonuclease domain interacts with nsp10 and provides part of the proofreading function that supports the high fidelity of the RNA polymerase complex (Robson et al., 2020). Several unusual features, such as the unusual zinc finger motifs, set it apart from other DEDD-type exonucleases (Chen et al., 2007), which are related to both nsp10 binding and catalytic activity. The MTase domain modifies the N7 of the guanosine cap of genomic and subgenomic viral RNAs, which is essential for the translation of viral proteins (Thoms et al., 2020). The location of this enzymatic activity within the RNA synthesis machinery ensures that newly synthesized RNA is rapidly capped and thus stabilized. As a strategy, we used constructs, which allow coexpression of both nsp14 and nsp10 (pRSFDuet and pETDuet, respectively). Production of isolated fl nsp14 was successful, however, with limited yield and stability (Table 2). Expression of the isolated MTase domain resulted in soluble protein with 27.5 kDa mass that was amenable to NMR characterization (Figure 2), although only under reducing conditions and in the presence of high (0.4 M) salt concentration.

nsp15

The poly-U-specific endoribonuclease nsp15 was one of the very first SCoV2 structures deposited in the PDB [6VWW, (Kim et al., 2020)]. Its function has been suggested to be related to the removal of U-rich RNA elements, preventing recognition by the innate immune system (Deng et al., 2017), even though the precise mechanism remains to be established. The exact role of the three domains (N-terminal, middle, and C-terminal catalytic domain) also remains to be characterized in more detail (Kim et al., 2020). Here, the sufficient yield of fl nsp15 during expression supported purification of pure protein, which, however, showed limited stability in solution (Table 2).

nsp16

The MTase reaction catalyzed by nsp16 is dependent on nsp10 as a cofactor (Krafcikova et al., 2020). In this reaction, the 2'-OH group of nucleotide +1 in genomic and subgenomic viral RNA is methylated, preventing recognition by the innate immune system. Since both nsp14 and nsp16 are in principle susceptible to inhibition by MTase inhibitors, a drug targeting both enzymes would be highly desirable (Bouvet et al., 2010). nsp16 is the last protein being encoded by ORF1ab, and only its N-terminus is formed by cleavage by the M^{PRO} nsp5. Employing a similar strategy to that for nsp14, nsp16 constructs were designed with the possibility of nsp10 coexpression. Expression of fl nsp16 resulted in good yields, when expressed both isolated and together with nsp10. The protein, however, is in either case



unstable in solution and highly dependent on reducing buffer conditions (Table 2). The purification procedures of nsp16 were adapted with minor modifications from a previous X-ray crystallography study (Rosas-Lemus et al., 2020).

Structural Proteins and Accessory ORFs

Besides establishing expression and purification protocols for the nsps, we also developed protocols and obtained pure mg quantities of the SCoV2 structural proteins E, M, and N, as well as literally all accessory proteins. With the exception of the relatively well-behaved nucleocapsid (N) protein, SCoV2 E, M,

and the remaining accessory proteins represent a class of mostly small and relatively poorly characterized proteins, mainly due to their difficult behavior in classical expression systems.

We used wheat-germ cell-free protein synthesis (WG-CFPS) for the successful production, solubilization, purification, and, in part, initial NMR spectroscopic investigation of ORF3a, ORF6, ORF7b, ORF8, ORF9b, and ORF14 accessory proteins, as well as E and M in mg quantities using the highly efficient translation machinery extracted from wheat-germs (Figures 6A–D).

ORF3a

The protein from ORF3a in SCoV2 corresponds to the accessory protein 3a in SCoV, with homology of more than 70% (Table 1). It has 275 amino acids, and its structure has recently been determined (Kern et al., 2020). The structure of SCoV2 3a displays a dimer, but it can also form higher oligomers. Each monomer has three TM helices and a cytosolic β -strand rich domain. SCoV2 ORF3a is a cation channel, and its structure has been solved by electron microscopy in nanodiscs. In SCoV, 3a is a structural component and was found in recombinant virus-like particles (Liu et al., 2014), but is not explicitly needed for their formation. The major challenge for NMR studies of this largest accessory protein is its size, independent of its employment in solid state or solution NMR spectroscopy.

As most other accessory proteins described in the following, ORF3a has been produced using WG-CFPS and was expressed in soluble form in the presence of Brij-58 (Figure 6C). It is copurified with a small heat-shock protein of the HSP20 family from the wheat-germ extract. The protocol described here is highly similar to that of the other cell-free synthesized accessory proteins. Where NMR spectra have been reported, the protein has been produced in a ^2H , ^{13}C , ^{15}N uniformly labeled form; otherwise, natural abundance amino acids were added to the reaction. The proteins were further affinity-purified in one step using Strep-Tactin resin, through the Strep-tag II fused to their N- or C-terminus. For membrane proteins, protein synthesis and also purification were done in the presence of detergent.

About half a milligram of pure protein was generally obtained per mL of extract, and up to 3 ml wheat-germ extract have been used to prepare NMR samples.

ORF3b

The ORF3b protein is a putative protein stemming from a short ORF (57 aa) with no homology to existing SCoV proteins (Chan et al., 2020). Indeed, ORF3b gene products of SCoV2 and SCoV are considerably different, with one of the distinguishing features being the presence of premature stop codons, resulting in the expression of a drastically shortened ORF3b protein (Konno et al., 2020). However, the SCoV2 nucleotide sequence after the stop codon shows a high similarity to the SCoV ORF3b. Different C-terminal truncations seem to play a role in the interferon-antagonistic activity of ORF3b (Konno et al., 2020). ORF3b is the only protein that, using WG-CFPS, was not synthesized at all; i.e., it was neither observed in the total cell-free reaction nor in supernatant or pellet. This might be due to the premature stop codon, which was not considered. Constructs of ORF3b thus need to be redesigned.

ORF4 (Envelope Protein, E)

The SCoV2 envelope (E) protein is a small (75 amino acids), integral membrane protein involved in several aspects of the virus' life cycle, such as assembly, budding, envelope formation, and pathogenicity, as recently reviewed in (Schoeman and Fielding, 2020). Structural models for SCoV (Surya et al.,

2018) and the TM helix of SCoV2 (Mandala et al., 2020) E have been established. The structural models show a pentamer with a TM helix. The C-terminal part is polar, with charged residues interleaved, and is positioned on the membrane surface in SCoV. E was produced in a similar manner to ORF3a, using the addition of detergent to the cell-free reaction.

ORF5 (Membrane Glycoprotein, M)

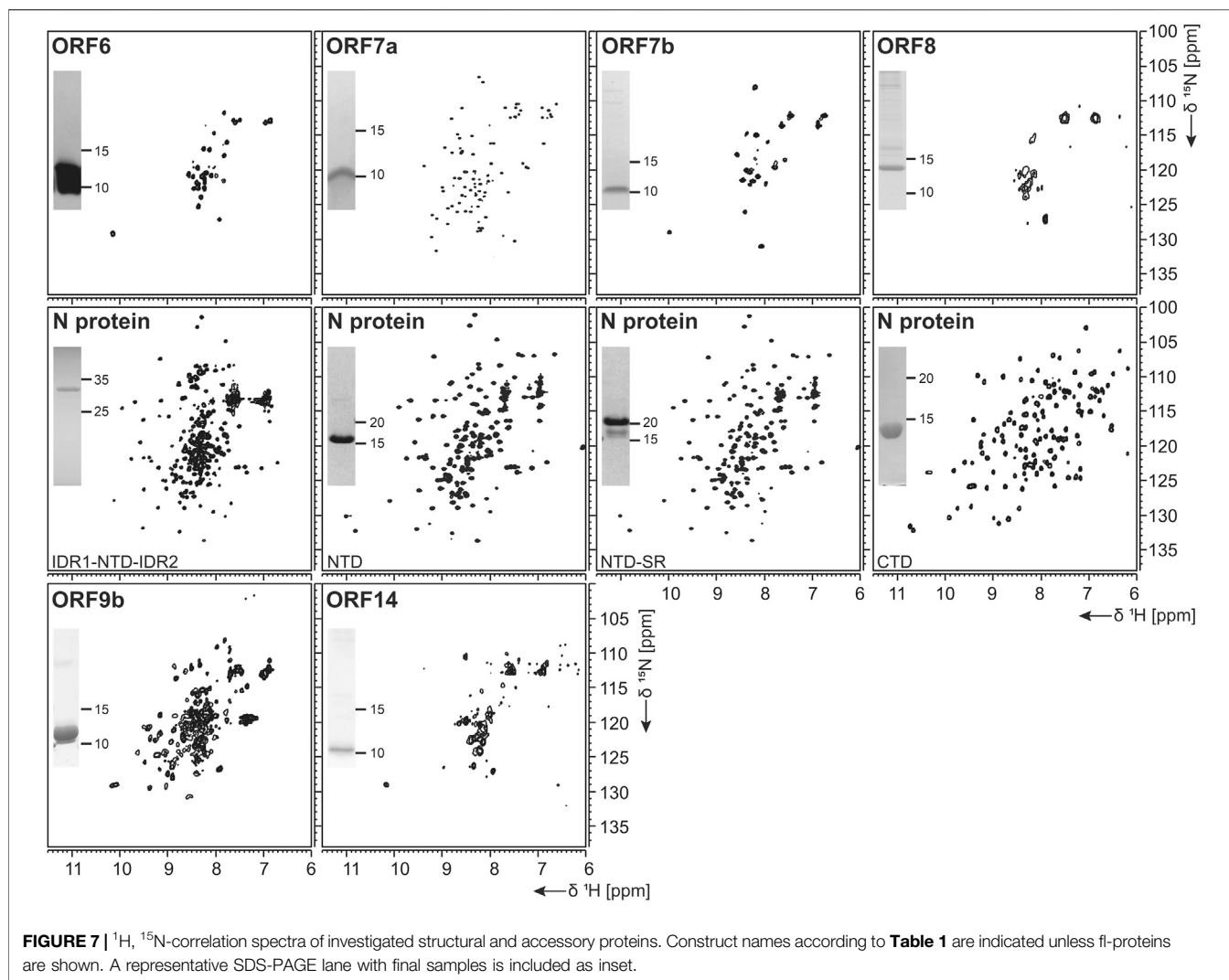
The M protein is the most abundant protein in the viral envelope and is believed to be responsible for maintaining the virion in its characteristic shape (Huang et al., 2004). M is a glycoprotein and sequence analyses predict three domains: A C-terminal endodomain, a TM domain with three predicted helices, and a short N-terminal ectodomain. M is essential for viral particle assembly. Intermolecular interactions with the other structural proteins, N and S to a lesser extent, but most importantly E (Vennema et al., 1996), seem to be central for virion envelope formation in coronaviruses, as M alone is not sufficient. Evidence has been presented that M could adopt two conformations, elongated and compact, and that the two forms fulfill different functions (Neuman et al., 2011). The lack of more detailed structural information is in part due to its small size, close association with the viral envelope, and a tendency to form insoluble aggregates when perturbed (Neuman et al., 2011). The M protein is readily produced using cell-free synthesis in the presence of detergent; as ORF3a, it is copurified with a small heat-shock protein of the HSP20 family (Figure 6B). Membrane-reconstitution will likely be necessary to study this protein.

ORF6

The ORF6 protein is incorporated into viral particles and is also released from cells (Huang et al., 2004). It is a small protein (61 aa), which has been found to concentrate at the endoplasmic reticulum and Golgi apparatus. In a murine coronavirus model, it was shown that expressing ORF6 increased virulence in mice (Zhao et al., 2009), and results indicate that ORF6 may serve an important role in the pathogenesis during SCoV infection (Liu et al., 2014). Also, it showed to inhibit the expression of certain STAT1-genes critical for the host immune response and could contribute to the immune evasion. ORF6 is expressed very well in WG-CFPS; the protein was fully soluble with detergents and partially soluble without them and was easily purified in the presence of detergent, but less efficiently in the absence thereof. Solution NMR spectra in the presence of detergent display narrow but few resonances, which correspond, in addition to the C-terminal STREP-tag, to the very C-terminal ORF6 protein residues.

ORF7a

SCoV2 protein 7a (121 aa) shows over 85% homology with the SCoV protein 7a. While the SCoV2 7a protein is produced and retained intracellularly, SCoV protein 7a has also been shown to be a structural protein incorporated into mature virions (Liu et al., 2014). 7a is one of the accessory proteins, of which a (partial) structure has been determined at high



resolution for SCoV2 (PDB 6W37). However, the very N-terminal signal peptide and the C-terminal membrane anchor, both highly hydrophobic, have not been determined experimentally yet.

Expression of the ORF7a ectodomain (ED) with a GB1 tag (Bogomolovas et al., 2009) was expected to produce reasonable yields. The IPRS purification resulted in a highly stable protein, as evidenced by the NMR data obtained (**Figure 7**).

ORF7b

Protein ORF7b is associated with viral particles in a SARS context (Liu et al., 2014). Protein 7b is one of the shortest ORFs with 43 residues. It shows a long hydrophobic stretch, which might correspond to a TM segment. It shows over 93% sequence homology with a bat coronavirus 7b protein (Liu et al., 2014). There, the cysteine residue in the C-terminal part is not conserved, which might facilitate structural studies. ORF7b has been synthesized successfully both from bacteria and by WG-CFPS in the presence of detergent and could be purified using a STREP-tag (**Table 2**). Due to the necessity of solubilizing agent

and its obvious tendency to oligomerize, structure determination, fragment screening, and interaction studies are challenging. However, we were able to record the first promising HSQC, as shown in **Figure 7**.

ORF8

ORF 8 is believed to be responsible for the evolution of *Betacoronaviruses* and their species jumps (Wu et al., 2016) and to have a role in repressing the host response (Tan et al., 2020). ORF 8 (121 aa) from SCoV2 does not apparently exist in SCoV on the protein level, despite the existence of a putative ORF. The sequences of the two homologs only show limited identity, with the exception of a small 7 aa segment, where, in SCoV, the glutamate is replaced with an aspartate. It, however, aligns very well with several coronaviruses endemic to animals, including Paguma and Bat (Chan et al., 2020). The protein comprises a hydrophobic peptide at its very N-terminus, likely corresponding to a signal peptide; the remaining part does not show any specific sequence features. Its

structure has been determined (PDB 7JTL) and shows a similar fold to ORF7a (Flower et al., 2020). In this study, ORF8 has been used both with (fl) and without signal peptide (Δ ORF8). We first tested the production of ORF8 in *E. coli*, but yields were low because of insolubility. Both ORF8 versions have then been synthesized in the cell-free system and were soluble in the presence of detergent. Solution NMR spectra, however, indicate that the protein is forming either oligomers or aggregates.

ORF9a (Nucleocapsid Protein, N)

The nucleocapsid protein (N) is important for viral genome packaging (Luo et al., 2006). The multifunctional RNA-binding protein plays a crucial role in the viral life cycle (Chang et al., 2014) and its domain architecture is highly conserved among coronaviruses. It comprises the N-terminal intrinsically disordered region (IDR1), the N-terminal RNA-binding globular domain (NTD), a central serine/arginine-(SR-) rich intrinsically disordered linker region (IDR2), the C-terminal dimerization domain (CTD), and a C-terminal intrinsically disordered region (IDR3) (Kang et al., 2020).

N represents a highly promising drug target. We thus focused our efforts not exclusively on the NTD and CTD alone, but, in addition, also provide protocols for IDR-containing constructs within the N-terminal part.

N-Terminal Domain

The NTD is the RNA-binding domain of the nucleocapsid (Kang et al., 2020). It is embedded within IDRs, functions of which have not yet been deciphered. Recent experimental and bioinformatic data indicate involvement in liquid-liquid phase separation (Chen et al., 2020a).

For the NTD, several constructs were designed, also considering the flanking IDRs (Table 1). In analogy to the available NMR [PDB 6YI3, (Dinesh et al., 2020)] and crystal [PDB 6M3M, (Kang et al., 2020)] structures of the SCoV2 NTD, boundaries for the NTD and the NTD-SR domains were designed to span residues 44–180 and 44–212, respectively. In addition, an extended IDR1-NTD-IDR2 (residues 1–248) construct was designed, including the N-terminal disordered region (IDR1), the NTD domain, and the central disordered linker (IDR2) that comprises the SR region. His-tagged NTD and NTD-SR were purified using IPRS and yielded approx. 3 mg/L in ^{15}N -labeled minimal medium. High protein quality and stability are supported by the available HSQC spectra (Figure 7).

The untagged IDR1-NTD-IDR2 was purified by IEC and yielded high amounts of ^{13}C , ^{15}N -labeled samples of 12 mg/L for further NMR investigations. The quality of our purification is confirmed by the available HSQC (Figure 7), and a near-complete backbone assignment of the two IDRs was achieved (Guseva et al., 2021; Schiavina et al., 2021). Notably, despite the structurally and dynamically heterogeneous nature of the N protein, the mentioned N constructs revealed a very good long-term stability, as shown in Table 2.

C-Terminal Domain

Multiple studies on the SCoV2 CTD, including recent crystal structures (Ye et al., 2020; Zhou et al., 2020), confirm the domain as dimeric. Its ability to self-associate seems to be necessary for viral replication and transcription (Luo et al., 2006). In addition, the CTD was shown to, presumably nonspecifically, bind ssRNA (Zhou et al., 2020).

Domain boundaries for the CTD were defined to comprise amino acids 247–364 (Table 1), in analogy to the NMR structure of the CTD from SCoV (PDB 2JW8, [Takeda et al., 2008]). Gene expression of His- or His-GST-tagged CTD yielded high amounts of soluble protein. Purification was achieved via IPRS. The CTD eluted as a dimer judged by its retention volume on the size-exclusion column and yielded good amounts (Table 2). The excellent protein quality and stability are supported by the available HSQC spectrum (Figure 7) and a near-complete backbone assignment (Korn et al., 2020b).

ORF9b

Protein 9b (97 aa) shows 73% sequence homology to the SCoV and also to bat virus (bat-SL-CoVZXC21) 9b protein (Chan et al., 2020). The structure of SCoV2 ORF9b has been determined at high resolution (PDB 6Z4U). Still, a significant portion of the structure was not found to be well ordered. The protein shows a β -sheet-rich structure and a hydrophobic tunnel, in which bound lipid was identified. How this might relate to membrane binding is not fully understood at this point. The differences in sequence between SCoV and SCoV2 are mainly located in the very N-terminus, which was not resolved in the structure (PDB 6Z4U). Another spot of deviating sequence not resolved in the structure is a solvent-exposed loop, which presents a potential interacting segment. ORF9b has been synthesized as a dimer (Figure 6E) using WG-CFPS in its soluble form. Spectra show a well-folded protein, and assignments are underway (Figure 6F).

ORF14 (ORF9c)

ORF14 (73 aa) remains, at this point in time, hypothetical. It shows 89% homology with a bat virus protein (bat-SL-CoVZXC21). It shows a highly hydrophobic part in its C-terminal region, comprising two negatively charged residues and a charged/polar N-terminus. The C-terminus is likely mediating membrane interaction. While ORF14 has been synthesized in the wheat-germ cell-free system in the presence of detergent and solution NMR spectra have been recorded, they hint at an aggregated protein (Figure 6E). Membrane-reconstitution of ORF14 revealed an unstable protein, which had been degraded during detergent removal.

ORF10

The ORF10 protein is comprised of 38 aa and is a hypothetical protein with unknown function (Yoshimoto, 2020). SCoV2 ORF10 displays 52.4% homology to SCoV ORF9b. The protein sequence is rich in hydrophobic residues, rendering expression and purification challenging. Expression of ORF10 as His-Trx-tagged or His-SUMO tagged fusion protein was possible; however, the ORF10 protein is poorly soluble and shows

partial unfolding, even as an uncleaved fusion protein. Analytical SEC hints at oligomerization under the current conditions.

DISCUSSION

The ongoing SCoV2 pandemic and its manifestation as the COVID-19 disease call for an urgent provision of therapeutics that will specifically target viral proteins and their interactions with each other and RNAs, which are crucial for viral propagation. Two “classical” viral targets have been addressed in comprehensive approaches soon after the outbreak in December 2019: the viral protease nsp5 and the RNA-dependent RNA polymerase (RdRp) nsp12. While the latter turned out to be a suitable target using the repurposed compound Remdesivir (Hillen et al., 2020), nsp5 is undergoing a broad structure-based screen against a battery of inhibitors in multiple places (Jin et al., 2020; Zhang et al., 2020), but with, as of yet, the limited outcome for effective medication. Hence, a comprehensive, reliable treatment of COVID-19 at any stage after the infection has remained unsuccessful.

Further viral protein targets will have to be taken into account in order to provide inhibitors with increased specificity and efficacy and preparative starting points for following potential generations of (SARS-)CoVs. Availability of those proteins in a recombinant, pure, homogenous, and stable form in milligrams is, therefore, a prerequisite for follow-up applications like vaccination, high-throughput screening campaigns, structure determination, and mapping of viral protein interaction networks. We here present, for the first time, a near-complete compendium of SCoV2 protein purification protocols that enable the production of large amounts of pure proteins.

The COVID19-NMR consortium was launched with the motivation of providing NMR assignments of all SCoV2 proteins and RNA elements, and enormous progress has been made since the outbreak of COVID-19 for both components [see Table 2 and (Wacker et al., 2020)]. Consequently, we have put our focus on producing proteins in stable isotope-labeled forms for NMR-based applications, e.g., the site-resolved mapping of interactions with compounds (Li and Kang, 2020). Relevant to a broad scientific community, we here report our protocols to suite perfectly any downstream biochemical or biomedical application.

Overall Success and Protein Coverage

As summarized in Table 2, we have successfully purified 80% of the SCoV2 proteins either in full or providing relevant fragments of the parent protein. Those include most of the nsps, where all of the known/predicted soluble domains have been addressed (Figure 1). For a very large part, we were able to obtain protein samples of high purity, homogeneity, and fold for NMR-based applications. We would like to point out a number of CoV proteins that, evidenced by their HSQCs, for the first time, provide access to structural information, e.g., the PL^{Pro} nsp3d and nsp3Y. Particularly for the nsp3 multidomain protein, we here present soluble samples of

almost the complete cytosolic region with more than 120 kDa in the form of excellent 2D NMR spectra (Figure 3), a major part of which fully backbone-assigned. We thus enable the exploitation of the largest and most enigmatic multifunctional SCoV2 protein through individual domains in solution, allowing us to study their concerted behavior with single residue resolution. Similarly, for nsp2, we provide a promising starting point for studying the so far neglected, often uncharacterized, and apparently unstructured proteins.

Driven by the fast-spreading COVID-19, we initially left out proteins that require advanced purification procedures (e.g., nsp12 and S) or where *a priori* information was limited (nsp4 and nsp6). This procedure seems justified with the time-saving approach of our effort in favor of the less attended proteins. However, we are in the process of collecting protocols for the missing proteins.

Different Complexities and Challenges

The compilation of protein production protocols, initially guided by information from CoV homologs (Table 1), has confronted us with very different levels of complexity. With some prior expectation toward this, we have shared forces to quickly “work off” the highly conserved soluble and small proteins and soon put focus into the processing of the challenging ones. The difficulties in studying this second class of proteins are due to their limited sequence conservation, no prior information, large molecular weights, insolubility, and so forth.

The nsp3e NAB represents one example where the available NMR structure of the SCoV homolog provided a *bona fide* template for selecting initial domain boundaries (Figure 4). The transfer of information derived from SCoV was straightforward; the transferability included the available protocol for the production of comparable protein amounts and quality, given the high sequence identity. In such cases, we found ourselves merely to adapt protocols and optimize yields based on slightly different expression vectors and *E. coli* strains.

However, in some cases, such transfer was unexpectedly not successful, e.g., for the short nsp1 GD. Despite intuitive domain boundaries with complete local sequence identity seen from the SCoV nsp1 NMR structure, it took considerable efforts to purify an analogous nsp1 construct, which is likely related to the impaired stability and solubility caused by a number of impacting amino acid exchanges within the domain’s flexible loops. In line with that, currently available structures of SCoV2 nsp1 have been obtained by crystallography or cryo-EM and include different buffers. As such, our initial design was insufficient in terms of taking into account the parameters mentioned above. However, one needs to consider those particular differences between the nsp1 homologs as one of the most promising target sites for potential drugs as they appear to be hotspots in the CoV evolution and will have essential effects for the molecular networks, both in the virus and with the host (Zust et al., 2007; Narayanan et al., 2015; Shen et al., 2019; Thoms et al., 2020).

A special focus was put on the production of the SCoV2 main protease nsp5, for which NMR-based screenings are ongoing. The

main protease is critical in terms of inhibitor design as it appears under constant selection, and novel mutants remarkably influence the structure and biochemistry of the protein (Cross et al., 2020). In the present study, the expression of the different constructs allowed us to characterize the protein in both its monomeric and dimeric forms. Comparison of NMR spectra reveals that the constructs with additional amino acids (GS and GHM mutant) display marked structural differences to the wild-type protein while being structurally similar among themselves (**Figure 5H**). The addition of two residues (GS) interferes with the dimerization interface, despite being similar to its native N-terminal amino acids (SGFR). We also introduced an active site mutation that replaces cysteine 145 with alanine (Hsu et al., 2005). Intriguingly, this active site mutation C145A, known to stabilize the dimerization of the main protease (Chang et al., 2007), supports dimer formation of the GS added construct (GS-nsp5 C145A) shown by its 2D NMR spectrum overlaying with the one of wild-type nsp5 (**Supplementary Table S14**). The NMR results are in line with SEC-MALS analyses (**Figure 5F**). Indeed, the additional amino acids at the N-terminus shift the dimerization equilibrium toward the monomer, whereas the mutation shifts it toward the dimer despite the N-terminal aa additions. This example underlines the need for a thorough and precise construct design and the detailed biochemical and NMR-based characterization of the final sample state. The presence of monomers vs. dimers will play an essential role in the inhibitor search against SCoV2 proteins, as exemplified by the particularly attractive nsp5 main protease target.

Exploiting Nonbacterial Expression

As a particular effort within this consortium, we included the so far neglected accessory proteins using a structural genomics procedure supported by wheat-germ cell-free protein synthesis. This approach allowed us previously to express a variety of difficult viral proteins in our hands (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Fogeron et al., 2017; Wang et al., 2019; Jirasko et al., 2020a). Within the workflow, we especially highlight the straightforward solubilization of the membrane proteins through the addition of detergent to the cell-free reaction, which allowed the production of soluble protein in milligram amounts compatible with NMR studies. While home-made extracts were used here, very similar extracts are available commercially (Cell-Free Sciences, Japan) and can thus be implemented by any lab without prior experience. Also, a major benefit of the WG-CFPS system for NMR studies lies in the high efficiency and selectivity of isotopic labeling. In contrast to cell-based expression systems, only the protein of interest is produced (Morita et al., 2003), which allows bypassing extensive purification steps. In fact, one-step affinity purification is in most cases sufficient, as shown for the different ORFs in this study. Samples could be produced for virtually all proteins, with the exception of the ORF3b construct used. With new recent insight into the stop codons present in this ORF, constructs will be adapted, which shall overcome the problems of ORF3b production (Konno et al., 2020).

For two ORFs, 7b and 8, we exploited a paralleled production strategy, i.e., both in bacteria and via cell-free synthesis. For those challenging proteins, we were, in principle, able to obtain pure samples from either expression system. However, for ORF7b, we found a strict dependency on detergents for follow-up work from both approaches. ORF8 showed significantly better solubility when produced in WG extracts compared to bacteria. This shows the necessity of parallel routes to take, in particular, for the understudied, biochemically nontrivial ORFs that might represent yet unexplored but highly specific targets to consider in the treatment of COVID-19.

Downstream structural analysis of ORFs produced with CFPS remains challenging but promising progress is being made in the light of SCoV2. Some solution NMR spectra show the expected number of signals with good resolution (e.g., ORF9b). As expected, however, most proteins cannot be straightforwardly analyzed by solution NMR in their current form, as they exhibit too large objects after insertion into micelles and/or by inherent oligomerization. Cell-free synthesized proteins can be inserted into membranes through reconstitution (Fogeron et al., 2015a; Fogeron et al., 2015b; Fogeron et al., 2016; Jirasko et al., 2020a; Jirasko et al., 2020b). Reconstitution will thus be the next step for many accessory proteins, but also for M and E, which were well produced by WG-CFPS. We will also exploit the straightforward deuteration in WG-CFPS (David et al., 2018; Wang et al., 2019; Jirasko et al., 2020a) that circumvents proton back-exchange, rendering denaturation and refolding steps obsolete (Tonelli et al., 2011). Nevertheless, the herein presented protocols for the production of non-nsp5 by WG-CFPS instantly enable their employment in binding studies and screening campaigns and thus provide a significant contribution to soon-to-come studies on SCoV2 proteins beyond the classical and convenient drug targets.

Altogether and judged by the ultimate need of exploiting recombinant SCoV2 proteins in vaccination and highly paralleled screening campaigns, we optimized sample amount, homogeneity, and long-term stability of samples. Our freely accessible protocols and accompanying NMR spectra now offer a great resource to be exploited for the unambiguous and reproducible production of SCoV2 proteins for the intended applications.

DATA AVAILABILITY STATEMENT

Assignments of backbone chemical shifts have been deposited at BMRB for proteins, as shown in **Table 2**, indicated by their respective BMRB IDs. All expression constructs are available as plasmids from <https://covid19-nmr.de/>.

AUTHOR CONTRIBUTIONS

NA, SK, NQ, MD, MN, ABö, HS, MH, and AS designed the study, compiled the protocols and NMR data, and wrote the manuscript. All authors contributed coordinative or practical work to the study. All authors contributed to the creation and collection of protein protocols and NMR spectra.

FUNDING

This work was supported by Goethe University (Corona funds), the DFG-funded CRC: “Molecular Principles of RNA-Based Regulation,” DFG infrastructure funds (project numbers: 277478796, 277479031, 392682309, 452632086, 70653611), the state of Hesse (BMRZ), the Fondazione CR Firenze (CERM), and the IWB-EFRE-program 20007375. This project has received funding from the European Union’s Horizon 2020 research and innovation program under Grant Agreement No. 871037. AS is supported by DFG Grant SCHL 2062/2-1 and by the JQYA at Goethe through project number 2019/AS01. Work in the lab of KV was supported by a CoRE grant from the University of New Hampshire. The FLI is a member of the Leibniz Association (WGL) and financially supported by the Federal Government of Germany and the State of Thuringia. Work in the lab of RM was supported by NIH (2R01EY021514) and NSF (DMR-2002837). BN-B was supported by the NSF GRFP. MC was supported by NIH (R25 GM055246 MBRS IMSD), and MS-P was supported by the HHMI Gilliam Fellowship. Work in the labs of KJ and KT was supported by Latvian Council of Science Grant No. VPP-COVID 2020/1-0014. Work in the UPAT’s lab was supported by the INSPIRED (MIS 5002550) project, which is implemented under the Action “Reinforcement of the Research and Innovation Infrastructure,” funded by the Operational Program “Competitiveness, Entrepreneurship and Innovation” (NSRF 2014–2020) and cofinanced by Greece and the EU (European Regional Development Fund) and the FP7 REGPOT CT-2011-285950–“SEE-DRUG” project (purchase of UPAT’s 700 MHz NMR equipment). Work in the CM-G lab was supported by the Helmholtz society. Work in the lab of ABö was supported by the CNRS, the French National Research Agency (ANR, NMR-SCoV2- ORF8), the Fondation de la Recherche Médicale (FRM, NMR-SCoV2-ORF8), and the IR-RMN-THC Fr3050 CNRS. Work in the lab of BM was supported by the Swiss National Science Foundation (Grant number 200020_188711), the Günthard Stiftung für Physikalische Chemie, and the ETH Zurich. Work in the labs of ABö and BM was supported by a common grant from SNF (grant 31CA30_196256). This work was supported by the ETH Zurich, the grant ETH 40 18 1, and the grant Krebsliga KFS 4903 08 2019. Work in the lab of the IBS Grenoble was supported by the Agence Nationale de Recherche (France)

REFERENCES

- Almeida, M. S., Johnson, M. A., Herrmann, T., Geralt, M., and Wüthrich, K. (2007). Novel beta-barrel fold in the nuclear magnetic resonance structure of the replicase nonstructural protein 1 from the severe acute respiratory syndrome coronavirus. *J. Virol.* 81 (7), 3151–3161. doi:10.1128/JVI.01939-06
- Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300 (5626), 1763–1767. doi:10.1126/science.1085658
- Bogomolovas, J., Simon, B., Sattler, M., and Stier, G. (2009). Screening of fusion partners for high yield expression and purification of bioactive viscotoxins. *Protein Expr. Purif.* 64 (1), 16–23. doi:10.1016/j.pep.2008.10.003

RA-COVID SARS2NUCLEOPROTEIN and European Research Council Advanced Grant DynamicAssemblies. Work in the CA lab was supported by Patto per il Sud della Regione Siciliana–CheMIST grant (CUP G77B17000110001). Part of this work used the platforms of the Grenoble Instruct-ERIC center (ISBG; UMS 3518 CNRS-CEA-UGA-EMBL) within the Grenoble Partnership for Structural Biology (PSB), supported by FRISBI (ANR-10-INBS-05-02) and GRAL, financed within the University Grenoble Alpes graduate school (Ecoles Universitaires de Recherche) CBH-EUR-GS (ANR-17-EURE-0003). Work at the UW-Madison was supported by grant numbers NSF MCB2031269 and NIH/NIAID AI123498. MM is a Ramón y Cajal Fellow of the Spanish AEI-Ministry of Science and Innovation (RYC2019-026574-I), and a “La Caixa” Foundation (ID 100010434) Junior Leader Fellow (LCR/BQ/PR19/11700003). Funded by project COV20/00764 from the Carlos III Institute of Health and the Spanish Ministry of Science and Innovation to MM and DVL. VDJ was supported by the Boehringer Ingelheim Fonds. Part of this work used the resources of the Italian Center of Instruct-ERIC at the CERM/CIRMMP infrastructure, supported by the Italian Ministry for University and Research (FOE funding). CF was supported by the Stiftung Polytechnische Gesellschaft. Work in the lab of JH was supported by NSF (RAPID 2030601) and NIH (R01GM123249).

ACKNOWLEDGMENTS

The authors thank Leonardo Gonnelli and Katharina Targaczewski for the valuable technical assistance. IBS acknowledges integration into the Interdisciplinary Research Institute of Grenoble (IRIG CEA). They acknowledge the Advanced Technologies Network Center of the University of Palermo to support infrastructures.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmolb.2021.653148/full#supplementary-material>

- Bojkova, D., Klann, K., Koch, B., Widera, M., Krause, D., Ciesek, S., et al. (2020). Proteomics of SARS-CoV-2-infected host cells reveals therapy targets. *Nature* 583 (7816), 469–472. doi:10.1038/s41586-020-2332-7
- Bouvet, M., Debarnot, C., Imbert, I., Selisko, B., Snijder, E. J., Canard, B., et al. (2010). *In vitro* reconstitution of SARS-coronavirus mRNA cap methylation. *PLoS Pathog.* 6 (4), e1000863. doi:10.1371/journal.ppat.1000863
- Cantini, F., Banci, L., Altincekic, N., Bains, J. K., Dhamotharan, K., Fuks, C., et al. (2020). (1)H, (13)C, and (15)N backbone chemical shift assignments of the apo and the ADP-ribose bound forms of the macrodomain of SARS-CoV-2 non-structural protein 3b. *Biomol. NMR Assign.* 14 (2), 339–346. doi:10.1007/s12104-020-09973-4
- Chan, J. F., Kok, K. H., Zhu, Z., Chu, H., To, K. K., Yuan, S., et al. (2020). Genomic characterization of the 2019 novel human-pathogenic coronavirus isolated from a patient with atypical pneumonia after visiting Wuhan. *Emerg. Microbes Infect.* 9 (1), 221–236. doi:10.1080/22221751.2020.1719902

- Chang, C. K., Hou, M. H., Chang, C. F., Hsiao, C. D., and Huang, T. H. (2014). The SARS coronavirus nucleocapsid protein--forms and functions. *Antivir. Res.* 103, 39–50. doi:10.1016/j.antiviral.2013.12.009
- Chang, H. P., Chou, C. Y., and Chang, G. G. (2007). Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. *Biophys. J.* 92 (4), 1374–1383. doi:10.1529/biophysj.106.091736
- Chen, H., Cui, Y., Han, X., Hu, W., Sun, M., Zhang, Y., et al. (2020a). Liquid-liquid phase separation by SARS-CoV-2 nucleocapsid protein and RNA. *Cell Res.* 30, 1143. doi:10.1038/s41422-020-00408-2
- Chen, J., Malone, B., Llewellyn, E., Grasso, M., Shelton, P. M. M., Olinares, P. D. B., et al. (2020b). Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex. *Cell* 182 (6), 1560–1573. doi:10.1016/j.cell.2020.07.033
- Chen, P., Jiang, M., Hu, T., Liu, Q., Chen, X. S., and Guo, D. (2007). Biochemical characterization of exoribonuclease encoded by SARS coronavirus. *J. Biochem. Mol. Biol.* 40 (5), 649–655. doi:10.5483/bmbrep.2007.40.5.649
- Chen, Y., Savinov, S. N., Mielech, A. M., Cao, T., Baker, S. C., and Mesecar, A. D. (2015). X-ray structural and functional studies of the three tandemly linked domains of non-structural protein 3 (nsp3) from murine hepatitis virus reveal conserved functions. *J. Biol. Chem.* 290 (42), 25293–25306. doi:10.1074/jbc.M115.662130
- Cornillez-Ty, C. T., Liao, L., Yates, J. R., 3rd, Kuhn, P., and Buchmeier, M. J. (2009). Severe acute respiratory syndrome coronavirus nonstructural protein 2 interacts with a host protein complex involved in mitochondrial biogenesis and intracellular signaling. *J. Virol.* 83 (19), 10314–10318. doi:10.1128/JVI.00842-09
- Cross, T. J., Takahashi, G. R., Diessner, E. M., Crosby, M. G., Farahmand, V., Zhuang, S., et al. (2020). Sequence characterization and molecular modeling of clinically relevant variants of the SARS-CoV-2 main protease. *Biochemistry* 59 (39), 3741–3756. doi:10.1021/acs.biochem.0c00462
- David, G., Fogeron, M. L., Schledorn, M., Montserret, R., Haselmann, U., Penzel, S., et al. (2018). Structural studies of self-assembled subviral particles: combining cell-free expression with 110 kHz MAS NMR spectroscopy. *Angew. Chem. Int. Ed. Engl.* 57 (17), 4787–4791. doi:10.1002/anie.201712091
- Davies, J. P., Almasy, K. M., McDonald, E. F., and Plate, L. (2020). Comparative multiplexed interactomics of SARS-CoV-2 and homologous coronavirus non-structural proteins identifies unique and shared host-cell dependencies. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.07.13.201517
- Deng, X., Hackbart, M., Mettelman, R. C., O'Brien, A., Mielech, A. M., Yi, G., et al. (2017). Coronavirus nonstructural protein 15 mediates evasion of dsRNA sensors and limits apoptosis in macrophages. *Proc. Natl. Acad. Sci. U.S.A.* 114 (21), E4251–E4260. doi:10.1073/pnas.1618310114
- Dinesh, D. C., Chalupska, D., Silhan, J., Koutna, E., Nencka, R., Veverka, V., et al. (2020). Structural basis of RNA recognition by the SARS-CoV-2 nucleocapsid phosphoprotein. *PLoS Pathog.* 16 (12), e1009100. doi:10.1371/journal.ppat.1009100
- Dudas, F. D., Puglisi, R., Korn, S. M., Alfano, C., Kelly, G., Monaca, E., et al. (2021). Backbone chemical shift spectral assignments of coronavirus-2 non-structural protein nsp9. *Biomol. NMR Assign.* 2021, 1–10. doi:10.1007/s12104-020-09992-1
- Egloff, M. P., Ferron, F., Campanacci, V., Longhi, S., Rancurel, C., Dutartre, H., et al. (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 (11), 3792–3796. doi:10.1073/pnas.0307877101
- Espósito, D., Mehalko, J., Drew, M., Snead, K., Wall, V., Taylor, T., et al. (2020). Optimizing high-yield production of SARS-CoV-2 soluble spike trimers for serology assays. *Protein Expr. Purif.* 174, 105686. doi:10.1016/j.pep.2020.105686
- Finkel, Y., Mizrahi, O., Nachshon, A., Weingarten-Gabbay, S., Morgenstern, D., Yahalom-Ronen, Y., et al. (2020). The coding capacity of SARS-CoV-2. *Nature* 589, 125. doi:10.1038/s41586-020-2739-1
- Flower, T. G., Buffalo, C. Z., Hooy, R. M., Allaire, M., Ren, X., and Hurley, J. H. (2020). Structure of SARS-CoV-2 ORF8, a rapidly evolving coronavirus protein implicated in immune evasion. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.08.27.270637
- Fogeron, M. L., Badillo, A., Jirasko, V., Gouttenoire, J., Paul, D., Lancien, L., et al. (2015a). Wheat germ cell-free expression: two detergents with a low critical micelle concentration allow for production of soluble HCV membrane proteins. *Protein Expr. Purif.* 105, 39–46. doi:10.1016/j.pep.2014.10.003
- Fogeron, M. L., Badillo, A., Penin, F., and Böckmann, A. (2017). Wheat germ cell-free overexpression for the production of membrane proteins. *Methods Mol. Biol.* 1635, 91–108. doi:10.1007/978-1-4939-7151-0_5
- Fogeron, M. L., Jirasko, V., Penzel, S., Paul, D., Montserret, R., Danis, C., et al. (2016). Cell-free expression, purification, and membrane reconstitution for NMR studies of the nonstructural protein 4B from hepatitis C virus. *J. Biomol. NMR* 65 (2), 87–98. doi:10.1007/s10858-016-0040-2
- Fogeron, M. L., Paul, D., Jirasko, V., Montserret, R., Lacabanne, D., Molle, J., et al. (2015b). Functional expression, purification, characterization, and membrane reconstitution of non-structural protein 2 from hepatitis C virus. *Protein Expr. Purif.* 116, 1–6. doi:10.1016/j.pep.2015.08.027
- Frick, D. N., Viridi, R. S., Vuksanovic, N., Dahal, N., and Silvaggi, N. R. (2020). Molecular basis for ADP-ribose binding to the Mac1 domain of SARS-CoV-2 nsp3. *Biochemistry* 59 (28), 2608–2615. doi:10.1021/acs.biochem.0c00309
- Gallo, A., Tsika, A. C., Fourkiotis, N. K., Cantini, F., Banci, L., Sreeramulu, S., et al. (2020). ¹H, ¹³C and ¹⁵N chemical shift assignments of the SUD domains of SARS-CoV-2 non-structural protein 3c: “the N-terminal domain-SUD-N”. *Biomol. NMR Assign.* 2020, 1–5. doi:10.1007/s12104-020-09987-y
- Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., et al. (2020). Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 368 (6492), 779–782. doi:10.1126/science.abb7498
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9
- Graham, R. L., Sims, A. C., Brockway, S. M., Baric, R. S., and Denison, M. R. (2005). The nsp2 replicase proteins of murine hepatitis virus and severe acute respiratory syndrome coronavirus are dispensable for viral replication. *J. Virol.* 79 (21), 13399–13411. doi:10.1128/JVI.79.21.13399-13411.2005
- Guseva, S., Perez, L. M., Camacho-Zarco, A., Bessa, L. M., Salvi, N., Malki, A., et al. (2021). (¹H), (¹³C) and (¹⁵N) Backbone chemical shift assignments of the n-terminal and central intrinsically disordered domains of SARS-CoV-2 nucleoprotein. *Biomol NMR Assign.* doi:10.1007/s12104-021-10014-x
- Hagemeyer, M. C., Monastyrska, I., Griffith, J., van der Sluijs, P., Voortman, J., van Bergen en Henegouwen, P. M., et al. (2014). Membrane rearrangements mediated by coronavirus nonstructural proteins 3 and 4. *Virology* 458–459, 125–135. doi:10.1016/j.virol.2014.04.027
- Hillen, H. S., Kokic, G., Farnung, L., Dienemann, C., Tegunov, D., and Cramer, P. (2020). Structure of replicating SARS-CoV-2 polymerase. *Nature* 584 (7819), 154–156. doi:10.1038/s41586-020-2368-8
- Hsu, M. F., Kuo, C. J., Chang, K. T., Chang, H. C., Chou, C. C., Ko, T. P., et al. (2005). Mechanism of the maturation process of SARS-CoV 3CL protease. *J. Biol. Chem.* 280 (35), 31257–31266. doi:10.1074/jbc.M502577200
- Huang, Y., Yang, Z. Y., Kong, W. P., and Nabel, G. J. (2004). Generation of synthetic severe acute respiratory syndrome coronavirus pseudoparticles: implications for assembly and vaccine production. *J. Virol.* 78 (22), 12557–12565. doi:10.1128/JVI.78.22.12557-12565.2004
- Hurst, K. R., Koetzner, C. A., and Masters, P. S. (2013). Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *J. Virol.* 87 (16), 9159–9172. doi:10.1128/JVI.01275-13
- Ishida, T., and Kinoshita, K. (2007). PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Res.* 35, W460–W464. doi:10.1093/nar/gkm363
- Jia, Z., Yan, L., Ren, Z., Wu, L., Wang, J., Guo, J., et al. (2019). Delicate structural coordination of the severe acute respiratory syndrome coronavirus Nsp13 upon ATP hydrolysis. *Nucleic Acids Res.* 47 (12), 6538–6550. doi:10.1093/nar/gkz409
- Jiang, H. W., Li, Y., Zhang, H. N., Wang, W., Yang, X., Qi, H., et al. (2020). SARS-CoV-2 proteome microarray for global profiling of COVID-19 specific IgG and IgM responses. *Nat. Commun.* 11 (1), 3581. doi:10.1038/s41467-020-17488-8
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582 (7811), 289–293. doi:10.1038/s41586-020-2223-y
- Jirasko, V., Lakomek, N. A., Penzel, S., Fogeron, M. L., Bartenschlager, R., Meier, B. H., et al. (2020a). Proton-detected solid-state NMR of the cell-free synthesized

- α -helical transmembrane protein NS4B from hepatitis C virus. *Chembiochem* 21 (10), 1453–1460. doi:10.1002/cbic.201900765
- Jirasko, V., Lends, A., Lakomek, N. A., Fogeron, M. L., Weber, M. E., Malär, A. A., et al. (2020b). Dimer organization of membrane-associated NS5A of hepatitis C virus as determined by highly sensitive 1 H-detected solid-state NMR. *Angew. Chem. Int. Ed.* 60 (10), 5339–5347. doi:10.1002/anie.202013296
- Johnson, M. A., Chatterjee, A., Neuman, B. W., and Wüthrich, K. (2010). SARS coronavirus unique domain: three-domain molecular architecture in solution and RNA binding. *J. Mol. Biol.* 400 (4), 724–742. doi:10.1016/j.jmb.2010.05.027
- Joseph, J. S., Saikatendu, K. S., Subramanian, V., Neuman, B. W., Brooun, A., Griffith, M., et al. (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 (16), 7894–7901. doi:10.1128/JVI.00467-06
- Kamitani, W., Narayanan, K., Huang, C., Lokugamage, K., Ikegami, T., Ito, N., et al. (2006). Severe acute respiratory syndrome coronavirus nsp1 protein suppresses host gene expression by promoting host mRNA degradation. *Proc. Natl. Acad. Sci. U.S.A.* 103 (34), 12885–12890. doi:10.1073/pnas.0603144103
- Kang, S., Yang, M., Hong, Z., Zhang, L., Huang, Z., Chen, X., et al. (2020). Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. *Acta Pharm. Sin. B* 10 (7), 1228–1238. doi:10.1016/j.apsb.2020.04.009
- Kern, D. M., Sorum, B., Mali, S. S., Hoel, C. M., Sridharan, S., Remis, J. P., et al. (2020). Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *bioRxiv* 17, 156554. doi:10.1101/2020.06.17.156554
- Khan, M. T., Zeb, M. T., Ahsan, H., Ahmed, A., Ali, A., Akhtar, K., et al. (2020). SARS-CoV-2 nucleocapsid and Nsp3 binding: an in silico study. *Arch. Microbiol.* 203, 59. doi:10.1007/s00203-020-01998-6
- Kim, Y., Jedrzejczak, R., Maltseva, N. I., Wilamowski, M., Endres, M., Godzik, A., et al. (2020). Crystal structure of Nsp15 endoribonuclease NendoU from SARS-CoV-2. *Protein Sci.* 29 (7), 1596–1605. doi:10.1002/pro.3873
- Kirchdoerfer, R. N., and Ward, A. B. (2019). Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nat. Commun.* 10 (1), 2342. doi:10.1038/s41467-019-10280-3
- Konkolova, E., Klima, M., Nencka, R., and Boura, E. (2020). Structural analysis of the putative SARS-CoV-2 primase complex. *J. Struct. Biol.* 211 (2), 107548. doi:10.1016/j.jsb.2020.107548
- Konno, Y., Kimura, I., Uriu, K., Fukushi, M., Irie, T., Koyanagi, Y., et al. (2020). SARS-CoV-2 ORF3b is a potent interferon antagonist whose activity is increased by a naturally occurring elongation variant. *Cell Rep.* 32 (12), 108185. doi:10.1016/j.celrep.2020.108185
- Korn, S. M., Dhamotharan, K., Fürtig, B., Hengesbach, M., Löhr, F., Qureshi, N. S., et al. (2020a). 1H, 13C, and 15N backbone chemical shift assignments of the nucleic acid-binding domain of SARS-CoV-2 non-structural protein 3e. *Biomol. NMR Assign.* 14 (2), 329–333. doi:10.1007/s12104-020-09971-6
- Korn, S. M., Lambert, R., Fürtig, B., Hengesbach, M., Löhr, F., Richter, C., et al. (2020b). 1H, 13C, and 15N backbone chemical shift assignments of the C-terminal dimerization domain of SARS-CoV-2 nucleocapsid protein. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09995-y
- Krafčikova, P., Silhan, J., Nencka, R., and Boura, E. (2020). Structural analysis of the SARS-CoV-2 methyltransferase complex involved in RNA cap creation bound to sinefungin. *Nat. Commun.* 11 (1), 3717. doi:10.1038/s41467-020-17495-9
- Kubatova, N., Qureshi, N. S., Altincekic, N., Abele, R., Bains, J. K., Ceylan, B., et al. (2020). 1H, 13C, and 15N backbone chemical shift assignments of coronavirus-2 non-structural protein Nsp10. *Biomol. NMR Assign.* 2020, 1–7. doi:10.1007/s12104-020-09984-1
- Kusov, Y., Tan, J., Alvarez, E., Enjuanes, L., and Hilgenfeld, R. (2015). A G-quadruplex-binding macrodomain within the “SARS-unique domain” is essential for the activity of the SARS-coronavirus replication-transcription complex. *Virology* 484, 313–322. doi:10.1016/j.virol.2015.06.016
- Leao, J. C., Gusmao, T. P. L., Zazar, A. M., Leao Filho, J. C., Barkokebas Santos de Faria, A., Morais Silva, I. H., et al. (2020). Coronaviridae-old friends, new enemy! *Oral Dis.* 2020, 13447. doi:10.1111/odi.13447
- Lei, J., Kusov, Y., and Hilgenfeld, R. (2018). Nsp3 of coronaviruses: structures and functions of a large multi-domain protein. *Antivir. Res.* 149, 58–74. doi:10.1016/j.antiviral.2017.11.001
- Li, Q., and Kang, C. (2020). A practical perspective on the roles of solution NMR spectroscopy in drug discovery. *Molecules* 25 (13), 2974. doi:10.3390/molecules25132974
- Littler, D. R., Gully, B. S., Colson, R. N., and Rossjohn, J. (2020). Crystal structure of the SARS-CoV-2 non-structural protein 9, Nsp9. *iScience* 23 (7), 101258. doi:10.1016/j.isci.2020.101258
- Liu, D. X., Fung, T. S., Chong, K. K., Shukla, A., and Hilgenfeld, R. (2014). Accessory proteins of SARS-CoV and other coronaviruses. *Antivir. Res.* 109, 97–109. doi:10.1016/j.antiviral.2014.06.013
- Luo, H., Chen, J., Chen, K., Shen, X., and Jiang, H. (2006). Carboxyl terminus of severe acute respiratory syndrome coronavirus nucleocapsid protein: self-association analysis and nucleic acid binding characterization. *Biochemistry* 45 (39), 11827–11835. doi:10.1021/bi0609319
- Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., et al. (2015). Structural basis and functional analysis of the SARS coronavirus nsp14-nsp10 complex. *Proc. Natl. Acad. Sci. U.S.A.* 112 (30), 9436–9441. doi:10.1073/pnas.1508686112
- Mandala, V. S., McKay, M. J., Shcherbakov, A. A., Dregni, A. J., Kolocouris, A., and Hong, M. (2020). Structure and drug binding of the SARS-CoV-2 envelope protein transmembrane domain in lipid bilayers. *Nat. Struct. Mol. Biol.* 27, 1202. doi:10.1038/s41594-020-00536-8
- Miknis, Z. J., Donaldson, E. F., Umland, T. C., Rimmer, R. A., Baric, R. S., and Schultz, L. W. (2009). Severe acute respiratory syndrome coronavirus nsp9 dimerization is essential for efficient viral growth. *J. Virol.* 83 (7), 3007–3018. doi:10.1128/JVI.01505-08
- Mompean, M., Trevino, M. A., and Laurents, D. V. (2020). Towards targeting the disordered SARS-CoV-2 nsp2 C-terminal region: partial structure and dampened mobility revealed by NMR spectroscopy. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.11.09.374173
- Morita, E. H., Sawasaki, T., Tanaka, R., Endo, Y., and Kohno, T. (2003). A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci.* 12 (6), 1216–1221. doi:10.1110/ps.0241203
- Narayanan, K., Huang, C., Lokugamage, K., Kamitani, W., Ikegami, T., Tseng, C. T., et al. (2008). Severe acute respiratory syndrome coronavirus nsp1 suppresses host gene expression, including that of type I interferon, in infected cells. *J. Virol.* 82 (9), 4471–4479. doi:10.1128/JVI.02472-07
- Narayanan, K., Ramirez, S. I., Lokugamage, K. G., and Makino, S. (2015). Coronavirus nonstructural protein 1: common and distinct functions in the regulation of host and viral gene expression. *Virus Res.* 202, 89–100. doi:10.1016/j.virusres.2014.11.019
- Nelson, C. W., Ardern, Z., Goldberg, T. L., Meng, C., Kuo, C. H., Ludwig, C., et al. (2020). Dynamically evolving novel overlapping gene as a factor in the SARS-CoV-2 pandemic. *Elife* 9, 59633. doi:10.7554/eLife.59633
- Netzer, W. J., and Hartl, F. U. (1997). Recombination of protein domains facilitated by co-translational folding in eukaryotes. *Nature* 388 (6640), 343–349. doi:10.1038/41024
- Neuman, B. W., Joseph, J. S., Saikatendu, K. S., Serrano, P., Chatterjee, A., Johnson, M. A., et al. (2008). Proteomics analysis unravels the functional repertoire of coronavirus nonstructural protein 3. *J. Virol.* 82 (11), 5279–5294. doi:10.1128/JVI.02631-07
- Neuman, B. W., Kiss, G., Kunding, A. H., Bhella, D., Baksh, M. F., Connelly, S., et al. (2011). A structural analysis of M protein in coronavirus assembly and morphology. *J. Struct. Biol.* 174 (1), 11–22. doi:10.1016/j.jsb.2010.11.021
- Neuman, B. W. (2016). Bioinformatics and functional analyses of coronavirus nonstructural proteins involved in the formation of replicative organelles. *Antivir. Res.* 135, 97–107. doi:10.1016/j.antiviral.2016.10.005
- Oostra, M., Hagemeijer, M. C., van Gent, M., Bekker, C. P., te Lintelo, E. G., Rottier, P. J., et al. (2008). Topology and membrane anchoring of the coronavirus replication complex: not all hydrophobic domains of nsp3 and nsp6 are membrane spanning. *J. Virol.* 82 (24), 12392–12405. doi:10.1128/JVI.01219-08
- Oostra, M., te Lintelo, E. G., Deijs, M., Verheije, M. H., Rottier, P. J., and de Haan, C. A. (2007). Localization and membrane topology of coronavirus nonstructural protein 4: involvement of the early secretory pathway in replication. *J. Virol.* 81 (22), 12323–12336. doi:10.1128/JVI.01506-07
- Pavesi, A. (2020). New insights into the evolutionary features of viral overlapping genes by discriminant analysis. *Virology* 546, 51–66. doi:10.1016/j.virol.2020.03.007

- Robertson, M. P., Igel, H., Baertsch, R., Haussler, D., Ares, M., Jr., and Scott, W. G. (2005). The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* 3 (1), e5. doi:10.1371/journal.pbio.0030005
- Robson, F., Khan, K. S., Le, T. K., Paris, C., Demirbag, S., Barfuss, P., et al. (2020). Coronavirus RNA proofreading: molecular basis and therapeutic targeting. *Mol. Cell* 79 (5), 710–727. doi:10.1016/j.molcel.2020.07.027
- Rosas-Lemus, M., Minasov, G., Shuvalova, L., Inniss, N. L., Kiryukhina, O., Wiersum, G., et al. (2020). The crystal structure of nsp10-nsp16 heterodimer from SARS-CoV-2 in complex with S-adenosylmethionine. *bioRxiv* [Epub ahead of print]. doi:10.1101/2020.04.17.047498
- Salvi, N., Bessa, L. M., Guseva, S., Camacho-Zarco, A., Maurin, D., Perez, L. M., et al. (2021). ¹H, ¹³C and ¹⁵N backbone chemical shift assignments of SARS-CoV-2 nsp3a. *Biomol. NMR Assign.* 2021, 1–4. doi:10.1007/s12104-020-10001-8
- Schoeman, D., and Fielding, B. C. (2020). Is there a link between the pathogenic human coronavirus envelope protein and immunopathology? A review of the literature. *Front. Microbiol.* 11, 2086. doi:10.3389/fmicb.2020.02086
- Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L. A., et al. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat. Struct. Mol. Biol.* 27 (10), 959–966. doi:10.1038/s41594-020-0511-8
- Serrano, P., Johnson, M. A., Almeida, M. S., Horst, R., Herrmann, T., Joseph, J. S., et al. (2007). Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *J. Virol.* 81 (21), 12049–12060. doi:10.1128/JVI.00969-07
- Serrano, P., Johnson, M. A., Chatterjee, A., Neuman, B. W., Joseph, J. S., Buchmeier, M. J., et al. (2009). Nuclear magnetic resonance structure of the nucleic acid-binding domain of severe acute respiratory syndrome coronavirus nonstructural protein 3. *J. Virol.* 83 (24), 12998–13008. doi:10.1128/JVI.01253-09
- Schiavina, M., Pontoriero, L., Uversky, V. N., Felli, I. C., and Pierattelli, R. (2021). The highly flexible disordered regions of the SARS-CoV-2 nucleocapsid N protein within the 1-248 residue construct: Sequence-specific resonance assignments through NMR. *Biomol NMR Assign.* [in press].
- Shen, Z., Wang, G., Yang, Y., Shi, J., Fang, L., Li, F., et al. (2019). A conserved region of nonstructural protein 1 from alphacoronaviruses inhibits host gene expression and is critical for viral virulence. *J. Biol. Chem.* 294 (37), 13606–13618. doi:10.1074/jbc.RA119.009713
- Shin, D., Mukherjee, R., Grewe, D., Bojkova, D., Baek, K., Bhattacharya, A., et al. (2020). Papain-like protease regulates SARS-CoV-2 viral spread and innate immunity. *Nature* 587 (7835), 657–662. doi:10.1038/s41586-020-2601-5
- Snijder, E. J., Bredenbeek, P. J., Dobbe, J. C., Thiel, V., Ziebuhr, J., Poon, L. L., et al. (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 (5), 991–1004. doi:10.1016/s0022-2836(03)00865-9
- Surya, W., Li, Y., and Torres, J. (2018). Structural model of the SARS coronavirus E channel in LMPG micelles. *Biochim. Biophys. Acta Biomembr.* 1860 (6), 1309–1317. doi:10.1016/j.bbame.2018.02.017
- Sutton, G., Fry, E., Carter, L., Sainsbury, S., Walter, T., Nettleship, J., et al. (2004). The nsp9 replicase protein of SARS-coronavirus, structure and functional insights. *Structure* 12 (2), 341–353. doi:10.1016/j.str.2004.01.016
- Takai, K., Sawasaki, T., and Endo, Y. (2010). Practical cell-free protein synthesis system using purified wheat embryos. *Nat. Protoc.* 5 (2), 227–238. doi:10.1038/nprot.2009.207
- Takeda, M., Chang, C. K., Ikeya, T., Güntert, P., Chang, Y. H., Hsu, Y. L., et al. (2008). Solution structure of the c-terminal dimerization domain of SARS coronavirus nucleocapsid protein solved by the SAIL-NMR method. *J. Mol. Biol.* 380 (4), 608–622. doi:10.1016/j.jmb.2007.11.093
- Tan, J., Vonnrhein, C., Smart, O. S., Bricogne, G., Bollati, M., Kusov, Y., et al. (2009). The SARS-unique domain (SUD) of SARS coronavirus contains two macrodomains that bind G-quadruplexes. *Plos Pathog.* 5 (5), e1000428. doi:10.1371/journal.ppat.1000428
- Tan, Y., Schneider, T., Leong, M., Aravind, L., and Zhang, D. (2020). Novel immunoglobulin domain proteins provide insights into evolution and pathogenesis of SARS-CoV-2-related viruses. *mBio* 11 (3). doi:10.1128/mBio.00760-20
- Thoms, M., Buschauer, R., Ameismeier, M., Koepke, L., Denk, T., Hirschenberger, M., et al. (2020). Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* 369 (6508), 1249–1255. doi:10.1126/science.abc8665
- Tonelli, M., Singarapu, K. K., Makino, S., Sahu, S. C., Matsubara, Y., Endo, Y., et al. (2011). Hydrogen exchange during cell-free incorporation of deuterated amino acids and an approach to its inhibition. *J. Biomol. NMR* 51 (4), 467–476. doi:10.1007/s10858-011-9575-4
- Tonelli, M., Rienstra, C., Anderson, T. K., Kirchdoerfer, R., and Henzler-Wildman, K. (2020). ¹H, ¹³C, and ¹⁵N backbone and side chain chemical shift assignments of the SARS-CoV-2 non-structural protein 7. *Biomol. NMR Assign.* 2020, 1–5. doi:10.1007/s12104-020-09985-0
- Tvarogová, J., Madhugiri, R., Bylapudi, G., Ferguson, L. J., Karl, N., and Ziebuhr, J. (2019). Identification and characterization of a human coronavirus 229E nonstructural protein 8-associated RNA 3'-terminal adenyllyltransferase activity. *J. Virol.* 93 (12), e00291–e00319. doi:10.1128/JVI.00291-19
- Ullrich, S., and Nitsche, C. (2020). The SARS-CoV-2 main protease as drug target. *Bioorg. Med. Chem. Lett.* 30 (17), 127377. doi:10.1016/j.bmcl.2020.127377
- Vennema, H., Godeke, G. J., Rossen, J. W., Voorhout, W. F., Horzinek, M. C., Opstelten, D. J., et al. (1996). Nucleocapsid-independent assembly of coronavirus-like particles by co-expression of viral envelope protein genes. *EMBO J.* 15 (8), 2020–2028. doi:10.1002/j.1460-2075.1996.tb00553.x
- Wacker, A., Weigand, J. E., Akabayov, S. R., Altincekic, N., Bains, J. K., Banijamali, E., et al. (2020). Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* 48, 12415. doi:10.1093/nar/gkaa1013
- Wang, Y., Kirkpatrick, J., Zur Lage, S., Korn, S. M., Neissner, K., Schwalbe, H., et al. (2021). (¹H), (¹³C), and (¹⁵N) backbone chemical-shift assignments of SARS-CoV-2 non-structural protein 1 (leader protein). *Biomol NMR Assign.* doi:10.1007/s12104-021-10019-6
- Wang, S., Fogeron, M. L., Schledorn, M., Dujardin, M., Penzel, S., Burdette, D., et al. (2019). Combining cell-free protein synthesis and NMR into a tool to study capsid assembly modulation. *Front. Mol. Biosci.* 6, 67. doi:10.3389/fmolb.2019.00067
- Wolff, G., Limpens, R. W. A. L., Zevenhoven-Dobbe, J. C., Laugks, U., Zheng, S., de Jong, A. W. M., et al. (2020). A molecular pore spans the double membrane of the coronavirus replication organelle. *Science* 369 (6509), 1395–1398. doi:10.1126/science.abd3629
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579 (7798), 265–269. doi:10.1038/s41586-020-2008-3
- Wu, Z., Yang, L., Ren, X., Zhang, J., Yang, F., Zhang, S., et al. (2016). ORF8-related genetic evidence for Chinese horseshoe bats as the source of human severe acute respiratory syndrome coronavirus. *J. Infect. Dis.* 213 (4), 579–583. doi:10.1093/infdis/jiv476
- Ye, Q., West, A. M. V., Silletti, S., and Corbett, K. D. (2020). Architecture and self-assembly of the SARS-CoV-2 nucleocapsid protein. *Protein Sci.* 29, 1890. doi:10.1002/pro.3909
- Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., et al. (2020). Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368 (6498), 1499–1504. doi:10.1126/science.abc1560
- Yoshimoto, F. K. (2020). The proteins of severe acute respiratory syndrome coronavirus-2 (SARS CoV-2 or n-COV19), the cause of COVID-19. *Protein J.* 39 (3), 198–216. doi:10.1007/s10930-020-09901-4
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science* 368 (6489), 409–412. doi:10.1126/science.abb3405
- Zhao, J., Falcón, A., Zhou, H., Netland, J., Enjuanes, L., Pérez Breña, P., et al. (2009). Severe acute respiratory syndrome coronavirus protein 6 is required for optimal replication. *J. Virol.* 83 (5), 2368–2373. doi:10.1128/JVI.02371-08
- Zhou, R., Zeng, R., Von Brunn, A., and Lei, J. (2020). Structural characterization of the C-terminal domain of SARS-CoV-2 nucleocapsid protein. *Mol. Biomed.* 1 (2), 1–11. doi:10.1186/s43556-020-00001-4
- Züst, R., Cervantes-Barragán, L., Kuri, T., Blakqori, G., Weber, F., Ludewig, B., et al. (2007). Coronavirus non-structural protein 1 is a major pathogenicity factor: implications for the rational design of coronavirus vaccines. *PLoS Pathog.* 3 (8), e109. doi:10.1371/journal.ppat.0030109

Conflict of Interest: CH was employed by Signals GmbH & Co. KG.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Altincekic, Korn, Qureshi, Dujardin, Ninot-Pedrosa, Abele, Abi Saad, Alfano, Almeida, Alshamleh, de Amorim, Anderson, Anobom, Anorma, Bains, Bax, Blackledge, Blechar, Böckmann, Brigandat, Bula, Bütikofer, Camacho-Zarco, Carlomagno, Caruso, Ceylan, Chaikuad, Chu, Cole, Crosby, de Jesus, Dharmotharan, Felli, Ferner, Fleischmann, Fogeron, Fourkiotis, Fuks, Fürtig, Gallo, Gande, Gerez, Ghosh, Gomes-Neto, Gorbatyuk, Guseva, Hacker, Häfner, Hao, Hargittay, Henzler-Wildman, Hoch, Hohmann, Hutchison, Jaudzems, Jović, Kaderli, Kalniņš, Kaņepe, Kirchdoerfer, Kirkpatrick, Knapp, Krishnathas, Kutz, zur

Lage, Lambertz, Lang, Laurents, Lecoq, Linhard, Löhr, Malki, Bessa, Martin, Matzel, Maurin, McNutt, Mebus-Antunes, Meier, Meiser, Mompeán, Monaca, Montserret, Mariño Perez, Moser, Muhle-Goll, Neves-Martins, Ni, Norton-Baker, Pierattelli, Pontoriero, Pustovalova, Ohlenschläger, Orts, Da Poian, Pyper, Richter, Riek, Rienstra, Robertson, Pinheiro, Sabbatella, Salvi, Saxena, Schulte, Schiavina, Schwalbe, Silber, Almeida, Sprague-Piercy, Spyroulias, Sreeramulu, Tants, Tārs, Torres, Töws, Treviño, Trucks, Tsika, Varga, Wang, Weber, Weigand, Wiedemann, Wirmer-Bartoschek, Wirtz Martin, Zehnder, Hengesbach and Schlundt. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

GLOSSARY

- aa** Amino acid
- BEST** Band-selective excitation short-transient
- BMRB** Biomagnetic resonance databank
- CFPS** Cell-free protein synthesis
- CoV** Coronavirus
- CTD** C-terminal domain
- DEDD** Asp-Glu-Glu-Asp
- DMS** Dimethylsulfate
- E** Envelope protein
- ED** Ectodomain
- fl** Full-length
- GB1** Protein G B1 domain
- GD** Globular domain
- GF** Gel filtration
- GST** Glutathione-S-transferase
- His** Hisx-tag
- HSP** Heat-shock protein
- HSQC** Heteronuclear single quantum coherence
- IDP** Intrinsically disordered protein
- IDR** Intrinsically disordered region
- IEC** Ion exchange chromatography
- IMAC** Immobilized metal ion affinity chromatography
- IPRS** IMAC-protease cleavage-reverse IMAC-SEC;
- M** Membrane protein
- MERS** Middle East Respiratory Syndrome
- MHV** Murine hepatitis virus
- M^{pro}** Main protease
- MTase** Methyltransferase
- N** Nucleocapsid protein
- NAB** Nucleic acid-binding domain
- nsp** Nonstructural protein
- NTD** N-terminal domain
- PL^{pro}** Papain-like protease
- RdRP** RNA-dependent RNA polymerase
- S** Spike protein
- SARS** Severe Acute Respiratory Syndrome
- SEC** Size-exclusion chromatography
- SUD** SARS unique domain
- SUMO** Small ubiquitin-related modifier
- TEV** Tobacco etch virus
- TM** Transmembrane
- TROSY** Transverse relaxation-optimized spectroscopy
- Trx** Thioredoxin
- Ubl** Ubiquitin-like domain
- Ulp1** Ubiquitin-like specific protease 1
- WG** Wheat-germ.