

SARS-CoV-2 RNA-Dependent RNA Polymerase Proteomics and Interactions with
RNA-Binding FUS Protein

By:

RAMSEY C MARTZ

A Thesis Submitted to The Honors College

In Partial Fulfillment of the Bachelors degree With Honors in

Biochemistry

UNIVERSITY OF ARIZONA

M A Y 2 0 2 1

Approved by:

Dr. Jacob C Schwartz

Department of Chemistry and Biochemistry

Abstract

The SARS-CoV-2 (COVID-19) pandemic will forever be a staple of the 2020s, with countries shutting down across the world and catalyzing revolutions in biochemical engineering. However, the virus itself is a distant echo of the SARS-CoV-1 (SARS) epidemic from the early 2000s - all three of which being of the same family and genus of betacoronaviruses. The trend of coronavirus outbreaks prompted the study of the similarities between SARS and COVID-19 while also investigating the proliferation of coronaviruses. This paper studied the changes in the RNA-dependent RNA polymerase (RdRp), used in betacoronaviruses for gene expression and RNA genome replication, between SARS and COVID-19. Furthermore, proteomic studies were performed between strains of COVID-19 from across the world. Along with proteomics, the effect of FUS protein, a transcription factor that recruits and promotes RNA Polymerase II activity, was observed in transcription experiments with COVID-19 RdRp. Overall, the comparison between SARS and COVID-19 RdRp demonstrates a highly conserved structure and sequence - the uneven distribution of amino acid changes also suggests regions of the RdRp are conserved to maintain function. There were few mutations in the RdRp among COVID-19 strains compared to other proteins signifying a low rate of nonfatal mutations occurring. Lastly, the transcription assays with RdRp and FUS showed limited promotion of RdRp transcription with the addition of FUS, which may offer clues to COVID-19 taking advantage of other cellular machinery for proliferation and gene expression.

Introduction

Coronaviruses are a particularly famous classification of viruses, as SARS-CoV-2, also called COVID-19, and a member of the coronavirus family, is responsible for the 2020 pandemic that has shut down the entire world for months¹. Interestingly, SARS-CoV-2 is not the only coronavirus to have caused a pandemic/epidemic. Other infamous examples are Severe Acute Respiratory Syndrome (SARS or SARS-CoV-1) and Middle East Respiratory Syndrome (MERS or MERS-CoV), both of which occurred in the past twenty years with fatality rates of 10% and greater than 30% respectively². Although both viruses have frighteningly high mortality rates, the number of SARS-CoV-1 and MERS cases (approximately 10,000 altogether) pale in comparison to the estimates of total SARS-CoV-2 cases, which hovers around 65 million as of December

2020 with 1.5 million deaths (around 2.3% mortality rate)^{3,4}.

The term ‘coronavirus’ refers to the *Coronaviridae* family, which falls under the *Nidovirales* order. All viruses that fall into the *Nidovirales* order have positive-sense, single-stranded RNA genomes (+ssRNA)⁵. This allows for the RNA genes of the viruses to be recognized and transcribed by host cell machinery without the need for a complementary strand to be transcribed beforehand. The *Coronaviridae* family is the largest, and most widely studied, virus family of the order, and divides into sub-families, *Torovirinae* and *Coronavirinae*⁶. *Coronaviridae* viruses have similar structures and proteins that form their membrane; the four major ones being Spike (S), Membrane (M), Envelope (E), and Nucleocapsid (N) proteins. *Coronavirinae* viruses, however, are given their name because of the shape of the M and S protein on the outer shell of the viruses, taking on a

crown ('corona') shape. The *Coronavirinae* sub-family is further divided into four major genera; alpha-; beta-; gamma-; and delta-CoV, which are the concern of coronaviruses known to infect birds, amphibians, and mammals⁷. The alpha- and beta-CoV genera are closely related as they both infect mammalian hosts, including bats and humans⁸. Of the four genera of *Coronavirinae*, SARS, MERS, and SARS-CoV-2 all belong to the beta-CoV genus^{9,10}.

Table 1 is an illustration of the *Coronaviridae* family and other families within the *Nidovirales* order.

Coronavirus genomes are on the larger end of the virus RNA genome length spectrum, averaging around 30,000

nucleotides¹¹. The S, M, E, and N genes have been identified in novel CoV, including SARS, MERS, and SARS-CoV-2, along with many genes that code for nonstructural proteins (nsps). nsps do not have their own function alone, but they can interact for form functional complexes¹². One complex, the RNA-Dependent RNA Polymerase (RdRp), consists of three nsps in beta-CoVs: nsp7, nsp8, and nsp12. nsp7 and nsp8 are accessories to nsp12, which holds the active site of RNA polymerization. The complex is responsible for genome replication and gene expression in coronaviruses and uses subgenomic transcription to transcribe the RNA¹³. **Figure 1** below is an image of the RdRp complex of SARS-CoV-2.

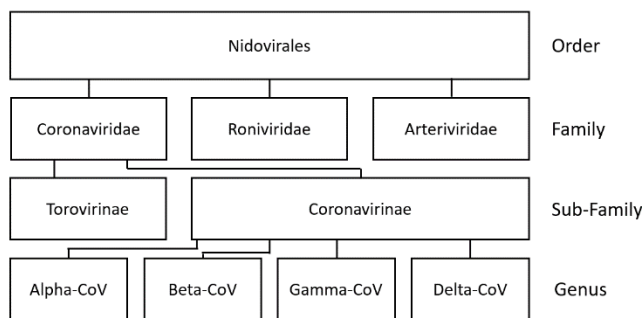


Table 1: Flowchart of *Nidovirales* order down to the genera of the *Coronavirinae* sub-family.

Being the largest protein in the RdRp complex, nsp12 also has several key domains that allow for RNA binding/stabilization. At the N-terminus of nsp12 is the N-terminus nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain,

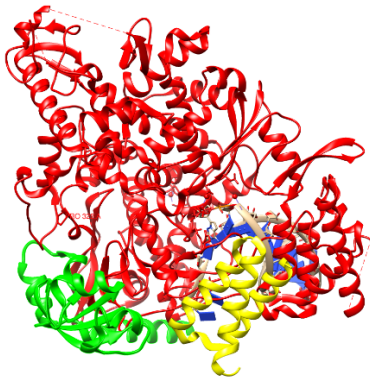


Figure 1: CoVID-19 RdRp protein structure from the PDB (7BV2)¹⁴, composed of nsp7 (yellow), nsp8 (green), and nsp12 (red).

which serves the function of transferring a phosphate group from ATP to nucleotide precursors that are later used in polymerizing nucleotides during transcription¹⁴. The finger, palm, and palm domains all interact with the template and transcribed RNA but in slightly different locations. The finger domains interact with the RNA template closest to the active site, the palm domain interacts with both the RNA template and product adjacent to the interaction sites between the RNA and finger domains, and the thumb domain interacts with the template and product (similar to the palm domain) but further away from the active site. Residues in

NH₃ NiRAN Interface Finger 1 Palm 1 Finger 2 Palm 2 Thumb COOH

Figure 2: An array of the nsp12 peptide sequence with the various domains throughout the protein. The finger and palm domains are segmented in the sequence, but the tertiary structure will bring the domains together.

the fingers and palm domains also help form the active site of RNA polymerization¹⁴.

Figure 2 illustrates the arrangement of the nsp12 domains along the amino acid sequence. SARS-CoV-2 can also have a second nsp8 subunit bind to nsp12 during transcription to stabilize the RNA product strands¹².

FET proteins refer to a group of RNA-binding proteins that play a role in transcription regulation¹⁵. For the scope of this paper, FUS is used to investigate the relationship between RNA-binding proteins and the SARS-CoV-2 RdRp complex. FUS protein has several Arg/Gly rich domains surrounding its RNA recognition motif (RRM), which allows for RNA binding¹⁶. FUS also has a low complexity (LC) domain that does not have a definite structure. FUS itself will have several monomers bind to an RNA sequence and polymerize with other

monomers to form a fiber. The fibers can then interact with the C-terminus of RNA Pol II to regulate its transcription¹⁷. Furthermore, mutations in FUS are involved in the pathology of some familial Amyotrophic Lateral Sclerosis (fALS) cases, and a translocation mutation with EWSR1 also leads to the development of Ewing's sarcoma^{18,19}.

Given RNA Pol II regulation via FUS-RNA binding and aggregation, there is a possibility of transcriptional regulation of SARS-CoV-2 RdRp regulation with FUS. The overall goal of these experiments is to first investigate genomic information about the RdRp complex and comparisons to another closely related coronavirus, SARS. After analyzing the proteomics of SARS-CoV-2 RdRp and variation among SARS-CoV-2 entries across the world, experiments were performed to observe any regulatory relationships between FUS and the RdRp complex through transcription assays.

Results

SARS-CoV-1/-2 RdRp Proteomics

Firstly, the RdRp complex structures of SARS-CoV-1 (SARS) and SARS-CoV-2 (SARS-CoV-2) were compared using structural data from the Protein Database^{14,15}.

Figure 3 is the superimposed structures of the RdRp complexes from both SARS-CoV-1 and SARS-CoV-2. Using the two structures, the root mean squared distance of the residue, tertiary, and quaternary structures of the complexes were examined. The root mean squared distance was to be 0.657 Å (65.7 pm) across 763 atoms. To give perspective, a chlorine molecule, Cl₂, has a radius of 1.75 Å (175 pm)¹⁹, making the average distance between atoms less than



Figure 3: CoVID-19 RdRp protein (7BV2, in pink)¹⁴ superimposed on the SARS-CoV-1 RdRp protein (6NUR, in blue)¹⁵. The extra appendage on SARS RdRp is an extra nsp7 subunit, which is commonly seen in both SARS and CoVID-19 complexes

half of diatomic chlorine. After comparing the tertiary structures of RdRp in SARS-CoV-1 and SARS-CoV-2, the amino acid sequence of both RdRp complexes were analyzed. Firstly, the amino acid sequences were compared and variations between the sequences were documented. **Table 2a** encompasses all the residue differences across nsp7 and -8, and **Table 2b** includes all differences in amino acids across the domains of nsp12. Amino acid variations are provided along with calculations of the percent identity (the proportion of amino acids that are the same between both sequences) and percent conservation (the proportion of amino acids that share chemical properties – polarity, acidity, and charge) of each nsp. The amino acid differences labeled

in red are amino acid differences that do not retain chemical properties.

Only one acid in nsp7 differed between SARS-CoV-1 and SARS-CoV-2 (99% identity); amino acid #70 is an arginine (a positively charged residue) in SARS-CoV-1 but is an asparagine (a neutral, yet polar, residue) in SARS-CoV-2. nsp8 similarly has very high percent conservation and identity with only five amino acids differing between SARS-CoV-1 and SARS-CoV-2. Most notably is that three mutations occur between residues 132 and 145 of both peptide sequences. The residue sequences in nsp12 between SARS-CoV-1 and SARS-CoV-2. Most notably is that three mutations occur between residues 132 and 145 of both peptide sequences.

2a

Protein (# of residues)	% Conservation	% Identity	Differences (SARS#COVID)
nsp7 (83)	99% (82/83)	99% (82/83)	R70N
nsp8 (197)	98% (195/197)	97% (192/197)	Y15F, V132I, G136N, N145T, N173S

2b

nsp12 Domain	% Conservation	% Identity	Differences (SARS#COVID)
NiRAN	96% (241/250)	93% (233/250)	S5Q, T6S, E40D, E62D, G63D, L66I, M77F, V90L, V98K, V106I, S185A, D198N, V224I, V226T, A227T, C230S, I234V
Interface	97% (192/197)	96% (190/197)	A252T, M257V, A259T, A262T, L265Y, C281K, I299V
Finger 1	100%	99% (181/182)	T611N
Palm 1	98% (44/45)	89% (40/45)	N643T, N647S, H679T, E680D, D683N
Finger 2	100%	100%	--
Palm 2	98% (122/124)	97% (120/124)	Y766F, N769T, A772S, A784S
Thumb	100%	100%	--

Table 2: 2a – table of conservation and identity of nsp7 and -8 between SARS and CoVID-19 and the amino acid differences between the two viruses, amino acids in red are non-conserved amino acids. Nsp8 demonstrated more residue changes between SARS and CoVID-19, likely due to its larger size. 2b – table of conservation and identity in nsp12 domains between SARS and CoVID-19. The NiRAN and palm domains had the most changes in amino acids while the finger and palm domains had the fewest (some having 100% identity).

The residue sequences in nsp12 between SARS-CoV-1 and SARS-CoV-2 demonstrated the most differences in amino acids, mostly attributed to the large sequence of nsp12 compared to nsp7 and -8. The NiRAN domain had the largest number of residue differences between SARS-CoV-1 and SARS-CoV-2, but the first palm domain had the highest proportion of residue changes (89% identity compared to the 93% identity of the NiRAN domain). The finger and thumb domains had the most retained sequences among the domains with only one amino acid differing across all three domains (which is a conserved amino acid in the first finger domain). It is also important to note that the finger and thumb domains interact with the RNA product polymerized from the RdRp complex. The interface domain also interacts with the RNA product and template, but the amino acids differences between SARS-CoV-1 and SARS-CoV-2 are found towards the N-terminal side of the residue

sequence, and the amino acids that differ are not involved with the interactions with RNA¹⁴. The most common conserved differences across the sequences involved valine – where valine in SARS-CoV-1 became another nonpolar amino acid in SARS-CoV-2 or a nonpolar amino acid in SARS-CoV-1 became valine. The most common non-conserved difference involved an alanine (nonpolar residue) in SARS-CoV-1 becoming a threonine (polar residue) in SARS-CoV-2, which was mostly seen in the NiRAN and interface domains.

SARS-CoV-2 Genome Analyses

After comparing the sequences of the RdRp complex proteins between SARS-CoV-1 and SARS-CoV-2, residue sequence analysis of the RdRp, along with the S and M proteins, was performed across SARS-CoV-2 sequence entries from across the world (locations of entries are listed in the Methods section). These comparisons used the genome entry from Wuhan, China as a

reference, and **Table 4** outlines the different mutations that occurred across SARS-CoV-2 strains and the frequency of the mutations. Similar to the previous tables, the residues in red are non-conserved mutations. The protein that experienced the most mutations across the strains was the S protein, which is essential to the binding of ACE2 receptors and the infection of cells²⁰. The most common mutation is amino acid #614, which converted an aspartic acid to a glycine, which is seen in 6 out of the 10 sequences analyzed. This mutation is much more common than the other S protein mutations, which are all only seen in 1 of the 10 sequences. The M protein, nsp7, and nsp8 all demonstrated no amino acid mutations across the 10 analyzed

sequences. However, nsp12 demonstrated a mutation that occurred in 7 of the 10 sequences, which is a non-conserved change in amino acid #322 from a proline (a nonpolar yet crucial residue in protein structure) to lysine (a positively charged amino acid). A structural analysis of the location of the amino acid in the tertiary structure of nsp12 shows that it is not close to any catalytic sites of the RdRp nor any regions that interact with nsp7 or -8, making this mutation unlikely to impact the catalytic activity of nsp12 and RdRp.

SARS-CoV-2 RdRp Purification

The protein growth protocols utilized *E. coli* bacteria, and the exact details of the

Protein	Mutations	Frequency
Spike	D614G, D138H, E554D, H49Y, S221W	6/10, 1/10, 1/10, 1/10, 1/10
Membrane	N/A	N/A
nsp7	N/A	N/A
nsp8	N/A	N/A
nsp12	P322L	7/10

Table 4: The list of mutations in SARS-CoV-2 proteins across 10 analyzed strains across the world. Spike (S) protein had the most mutations across the strains compared to the no mutations in the membrane (M) protein, nsp7, and -8. nsp12 had only one mutation present in seven of the 10 strains analyzed, but the mutation was not in a catalytically active region of the protein.

methodology are outlined in the Methods section. The elution of the RdRp complex proteins in the subtraction column occurred within a 40 mL window for all three nsps in the complex (chromatogram is shown in **Figure 4** along with the densitometry gel of the protein elution). The most striking aspect of the elution behavior of nsp7, -8, and -12 is the large range of weights between the three proteins. The densitometry gel of the elution seen in **Figure 4** showcases the large weight of the nsp12 protein, hovering around 100 kDa, and the small size of nsp7, found between 5- 10 kDa on the gel. nsp8 appears at the 20 kDa point, and the nsp8 band is much more saturated than nsp7 and -12 bands signifying a much higher concentration of

nsp8. Overall, the densitometry gel indicates that nsp8 was expressed at a much higher level than nsp7 and -12. The gel also showed that nsp7 was the least expressed protein of the three and is not visible in the gel when less than 3.0 μL of the protein solution was added to the lane. nsp12 had a darker band than nsp7 (the nsp12 band was visible when 1.0 μL of protein solution is added) but was not expressed to the same level as nsp8.

After further purifying the protein solution, the concentration of the solution was found to be approximately 7.8 mg/mL (corresponding to 9.4 μM). However, future purification would only yield around 1 or 2 mg/mL, which would often have little or no activity.

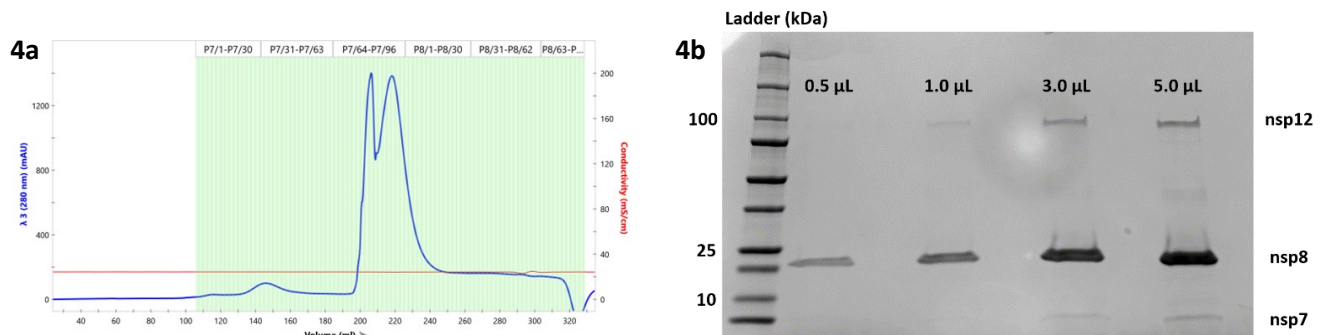


Figure 4: *Figure 4a* – elution chromatogram of the nsp mixture through the subtraction column, most of the target protein eluted out between the volume 200 to 240 mL. *Figure 4b* – densitometry of the collected fractions with variable volumes of protein added, the most highly expressed nsp was nsp8 with nsp12 and nsp7 being visually at lower concentrations.

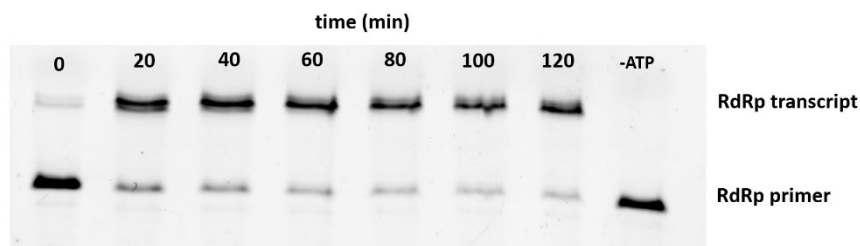


Figure 5: Time-course gel for SARS-CoV-2 RdRp transcription. The optimal time for future transcription assay was determined to be 60 minutes – due the amount of RNA transcript remaining constant after 1 hour of transcription.

RdRp Transcription Assay Development

A time-course transcription assay for the SARS-CoV-2 RdRp was performed to determine the time needed for majority of the RNA primer (template) to be polymerized into the transcript. **Figure 5** below includes the gel image for the transcription time-course assay including 2.4 μM of the RdRp solution, which demonstrates at least half of the primer being converted to the transcript after 20 minutes of transcribing the RNA. The amount of transcript plateaus after 60 minutes, which made 1 hour the optimal time for the transcription assays.

The 1-hour transcription assay time was incorporated into the transcription

protocol, and an optimal RdRp concentration is determined. ‘Optimal’ concentration is defined as having approximately 50% of the primer converted into the transcript after the 60-minute transcription assay. **Figure 6** is the gel of the RdRp titration into the transcription assay with concentration of RdRp ranging from 0.42 to 2.4 μM protein. The assays showcased that 1.2 μM RdRp converted about half of the RNA primer into the transcript product, making this the ideal concentration to use in transcription assays in future experiments.

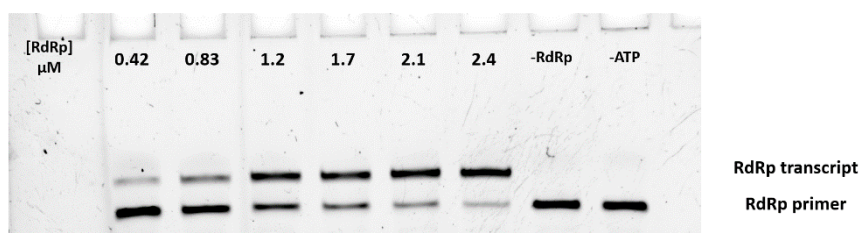


Figure 6: The RdRp titration-transcription assay gel. The optimal concentration for future experiments was determined to be 1.2 μM because this concentration lead to the conversion of 50% of the RNA primer into the polymerized transcript

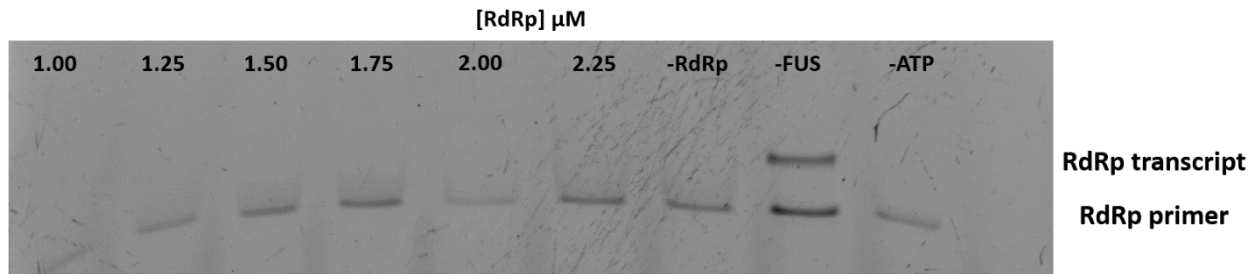


Figure 7: The first RdRp transcription assay titrating FUS into the reactions. The gel image above shows that there was no transcription activity in the reactions with FUS and RdRp, and the lane with no FUS added displayed normal RdRp transcription activity.

Effects of FUS on RdRp Transcription

Transcription assays running for one hour with 1.2 μM RdRp solution were titrated with FUS protein, and the transcription activity was analyzed. Two separate experiments were performed with a different range of FUS concentrations; one experiment had a concentration range from 1.00 to 2.25 μM FUS (**Figure 7**) and the other had a range from 0.125 to 4.00 μM FUS (**Figure 8**).

The first experiment demonstrates a complete loss of RdRp transcription activity with FUS added to the reactions. However, the sample without FUS (the “-FUS” lane)

had transcriptional activity return to the same level as the experiment shown in **Figure 6**.

The second experiment was performed to verify the loss of transcriptional activity seen with the addition of FUS while also expanding the range of FUS concentrations.

Figure 8 is the gel of the experiment performed and showcases a completely different effect on transcription activity.

Transcription appeared to be promoted with the addition of a low amount of FUS, which is evident in the darker transcript band of the 0.25 μM FUS lane. The 0.125 μM FUS and -FUS lanes appeared to have some

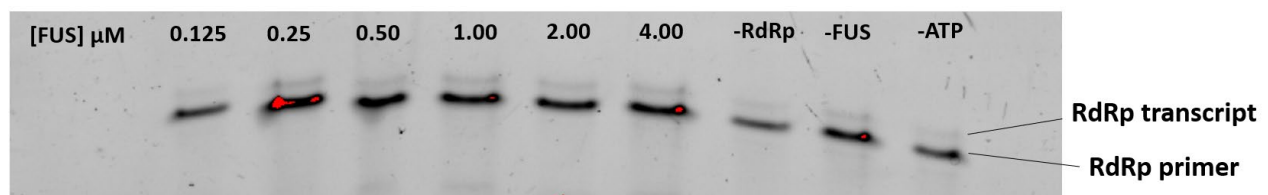


Figure 8: Second transcription assay experiment with RdRp and titrated FUS. This experiment used a later purification of nsp5, which had much lower transcriptional activity. However, the addition of FUS did not eliminate transcriptional activity, and instead promoted activity until the concentration of 0.25 μM FUS.

transcription but at a lower level than the 0.25 μM lane (which contradicts the higher level of transcription seen in the -FUS lane of **Figure 7**). Another similar experiment was performed by a colleague that yielded similar results, with the addition of FUS promoting transcriptional activity. However, the promotion of RdRp transcription activity plateaued when more than 0.25 μM FUS was added to the reaction. A graph quantifying the results of the repeated colleague experiment is shown in **Figure 9**. The notable increase in the transcription activity seen in FUS addition made the -FUS lane a statistical outlier (since it is the only lane that had less than 40% of the template transcribed into the product).

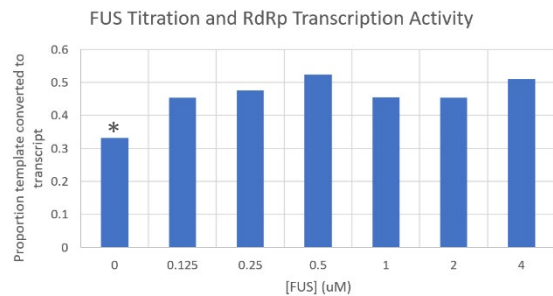


Figure 9: Graph of the colleague RdRp transcription assay with FUS titration. The addition of FUS appeared to increase the transcription activity, where all the reactions had more than 40% of the template converted to transcript. The reaction without FUS had lower transcription activity.

Discussion

The conserved sequence and structure of the RdRp in both SARS-CoV-1 and SARS-CoV-2 showcases the similar ancestry of both viruses, especially with SARS-CoV-1 and SARS-CoV-2 both being members of the betacoronavirus genus. The most common mutation between alanine and threonine can be explained by the RNA codons for each amino acid, where the four codons that code for alanine and threonine only differ in the first nucleotide in the codons. Alanine is coded by GCA, GCC, GCG, and GCU, whereas threonine is coded by ACA, ACC, ACG, and ACU. Not only do the codes differ by a single nucleotide, but the mutations were also transition mutations (converting a purine, adenine, into another purine, guanine).

The amino acid differences between SARS-CoV-1 and SARS-CoV-2 in nsp7 and -8 were minimal compared to nsp12, but their locations in the residue sequence and

structure were not in regions that interacted with nsp12 or the RNA bound and polymerized in the complex. Similarly, the nsp12 mutations were unevenly distributed, mostly being found in the NiRAN, interface, and palm domains. Although NiRAN is the domain that includes the active site for transferring phosphate groups from ATP to nucleotide precursors, the amino acid differences between the two viruses were not directly involved with the catalytic activity of the NiRAN domain. The interface domain also had amino acid differences not in the same region as RNA interactions. The finger and thumb domains were highly conserved between the SARS-CoV-1 and SARS-CoV-2 sequences, likely due to the fact these domains have more amino acids interacting with the RNA template and product to stabilize the strands. The one difference in the finger 1 domain involved a conserved amino acid change.

Mutations between the SARS-CoV-2 protein strains across the world were most prevalent in the S protein, which is responsible for the anchoring the virus to ACE2 receptors and allowing transport into cells²¹. The S protein had five mutations across 10 SARS-CoV-2 sequences, four of which only occurred in one sequence and one mutation appearing in six of the sequences. The large number of mutations, especially compared to the one mutation seen in nsp12 and zero mutations in the M protein, nsp7, and -8, indicates that the protein has a higher susceptibility to mutation than other proteins in SARS-CoV-2. However, it is important to acknowledge that survivorship bias comes into play, since it is entirely likely other proteins experience mutations at a similar frequency, but the mutations are fatal. It would be more accurate to say instead that the S protein has a higher rate of producing nonfatal mutations during proliferation. The mutations in the S protein are also not in the

RBM domain, which is responsible for ACE2 binding²¹.

The single mutation in nsp12 across SARS-CoV-2 strains mutated a proline residue to a lysine found in the interface domain. A structural analysis of the location of the residue demonstrates that the residue is found at the end of an α -helix is not involved in interactions with RNA or other amino acids. The original proline is also indicative of the structural, rather than catalytic, purpose of the amino acid. However, the use of lysine did not compromise the structural nor functional integrity of nsp12 or the RdRp.

The purification of nsp7, -8, and -12 revealed that nsp8 was expressed to a higher degree than nsp7 or -12. With nsp7 being the least expressed protein of the three, it is likely nsp7 was the limiting reagent in the formation of the complex. Research on SARS-CoV-1 has shown that the stoichiometry of the RdRp is not exactly 1:1:1 for the three nsps, but the formation of

the complex is 1:2:1 with 1 nsp7 and 2 nsp8 subunits bound to nsp12. The high demand for nsp8 may lead to its high expression seen during purification. However, the nsp8 would likely be in excess compared to nsp7 and -12, meaning there is likely a large concentration of free nsp8 subunits in solution.

It is likely few fully assembled complexes are in solution given the low concentration of nsp7, and there are nsp8-nsp12 complexes in solution without nsp7. Studies have demonstrated both nsp7 and -8 are necessary in the catalytic activity of nsp12, but future studies can look into the polymerase activity of nsp8-nsp12 complexes. Studies have documented the existence of a nsp7-nsp8 hexadecamer in SARS, which plays a role in RNA strand stabilization²². Given the conservation of sequences between SARS-CoV-1 and SARS-CoV-2, it is likely SARS-CoV-2 hexadecamers can form to stabilize RNA product, and the low concentrations of nsp7

may prevent the formation of hexadecamers and lead to compromised stability of RNA product.

The high concentration of nsp8 may also cause the protein concentration to appear higher than the actual concentration of catalytically active complexes. The low and nonexistent activity of later purifications of the nsps may be the result of low nsp12 and nsp7 concentrations in the solutions. If nsp8-nsp12 complexes demonstrate activity, it may provide some activity but not enough to rescue the low concentration of nsp12 during purification.

The initial transcription assay experiments with FUS and RdRp showcased no activity in lanes with both FUS and RdRp, yet the lane with only RdRp had full activity. There were two possibilities as to why there was no activity in the experiment: either the solution that the FUS was stored in contained salts or a contaminant that destroyed the RdRp complexes and their activity or FUS

regulates RdRp transcription. The FUS buffer and storage solution contain potassium chloride (KCl) and urea, which are known to denature protein at high concentrations and may result in the failure to transcribe RNA in the experiment. However, the experiments would have to be repeated to validate these results.

The repeated experiment was performed with a later purification of RdRp with reduced activity, as seen in the decreased activity of the no FUS lane. However, the repeated experiment demonstrated preserved RdRp transcription activity, but optimal activity was achieved with a small amount of FUS (0.25 μ M). The increased transcriptional activity is seen in higher concentrations of FUS, but the activity is only increased to a certain point. This rules out both the KCl/urea denaturation and the inhibitory FUS effect hypotheses, but there must have been a contaminant or experimental error that destroyed the

transcriptional activity of RdRp in the first experiment.

The colleague experiment that mirrored the protocol and design of the second experiment also supports this notion, and the graphed results of the transcription assay also supports the idea of limited FUS promotion of RdRp transcription. FUS is known to promote transcription by recruiting RNA Pol II to a transcription site, which may have a similar effect on RdRp during transcription reactions. This model suggests that the proliferation of SARS-CoV-2 may utilize transcription factors to promote the activity of RdRp. Studies have previously shown that coronaviruses can utilize transcription factors to promote proliferation while influencing intracellular processes, including apoptosis²². Further studies would have to investigate physical phenomena that occur between SARS-CoV-2 and FUS. Research has already shown that liquid-liquid phase separation occurs with high

concentrations of FUS, and the N protein of SARS-CoV-2 also participates in liquid-liquid phase separation^{15,23}. The transcription assay performed should also be repeated with varying conditions as well to investigate changes in RdRp activity under physiological stress conditions.

Methods

Proteomic Analysis

All amino acid sequence data for SARS-CoV-1 and SARS-CoV-2 (along with SARS-CoV-2 variants) were pulled from NCBI GenBank respectively. Amino acid sequences are compared through ProteinBlast on the National Center for Biotechnology Information website. The following list outlines the GenBank codes for each SARS-CoV-2 RdRp amino acid sequence used in the proteomics comparisons:

- Wuhan, China: NC_045512²⁴
- Virginia, USA: MT558707*
- Maricopa Country, AZ: MT339039²⁵
- USA (unspecific): MW594033*
- Turkey: MT675956*
- Lazio, Italy: MT525950²⁶
- Germany: MT845877*
- Hong Kong: MT114419*
- Israel: MT276597*
- South Korea: MT039890*

Protein Structure Analysis

All protein structure data for SARS-CoV-1 and SARS-CoV-2 were pulled from the RCSB Protein Database. All protein structures and figures were generated using Chimera. The following list outlines the PDB codes for the protein structures:

- SARS-CoV-1 RdRp: 6NUR¹⁵
- SARS-CoV-2 RdRp: 7BV2¹⁴

SARS-CoV-2 RdRp Growth & Purification

* - These SARS-CoV-2 genomes were direct submissions to the NCBI GenBank without an associated paper published

Protocols for RdRp growth was shared courtesy of the Peti lab. The plasmid vector used to growth the individual nsps was order from Ginkgo Bioworks, pGBW-m4046802 #145683. **Figure 10** below is a map of the vector, where the nsps have HisTags at the end of the peptide sequence.

The plasmid was introduced to *E. coli* bacteria cells for transformation. After growing, the cells were then homogenized using lysis buffer. All cell debris solution was filtered through a 0.22 µm syringe filter. The solution is then filtered through a HisTrap column to bind the crude nsp with the HisTags.

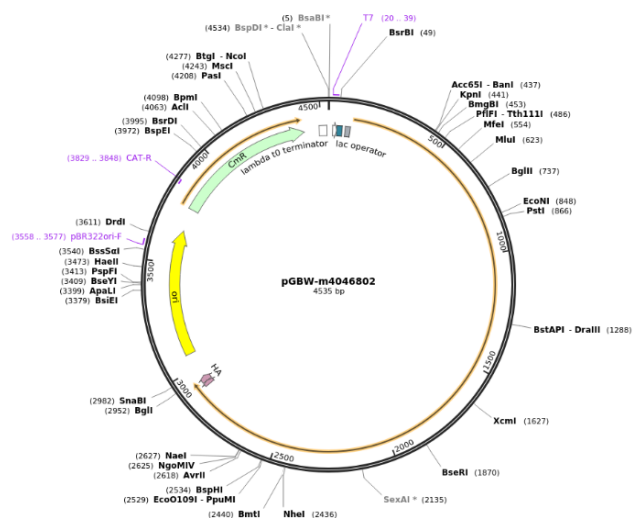


Figure 10: Map of the vector used to produce the RdRp nsp proteins with the additional HisTags (entry pGBW-m4046802 #145685 from Ginkgo Bioworks).

The protein solution is loaded into the column and then washed with a solution with a low concentration of imidazole (5 mM), which then creates a gradient that slowly increases imidazole concentration to 500 mM to knock the nsp off the column. Fractions with the highest absorbance, indicating nsp presence, were collected for further purification (as seen in **Figure 4a**). Protein densitometry is collected of the fractions and the crude protein concentrations are determined (as seen in **Figure 4b**).

With the concentrations of the samples, the HisTags can be cleaved off the proteins. For every 6 mg of protein yielded, 1 mL of TEV cleavage protein is added to the solution. The crude protein solution with added TEV is then placed into a dialysis tube to filter out imidazole and other ions overnight. The next day, the cleaved protein is further purified.

A 'subtraction column' is ran using the same running buffer with imidazole in the

previous HisTrap column to capture any proteins that bound to the column without HisTags. The cleaved nsp proteins will not bind to the column and will freely flow out as low concentrations of imidazole is introduced. Fractions with highest absorbance are collected to be concentrated. After concentrating the samples, the final protein concentration is determined, and the samples are aliquoted for transcription assay use.

Flag-FUS (FF) RNA Growth & Purification

The isolated FF RNA and DNA is then treated with DNase I and incubated at 37° C for 1 hour to destroy any remaining DNA. 3 M Sodium Acetate (pH = 5.2) is added to lightly acidify the solution and stabilize the RNA. An equal volume of 1:1 phenol:chloroform solution is added and the aqueous layer is extracted and washed two more times with equal volumes of pure chloroform. Two volumes of pure ethanol are added to the aqueous solution and then

incubated at -80°C for 30 minutes to precipitate out the FF RNA.

The mixture is then centrifuged at 4°C for 10 minutes, and the ethanol supernatant is removed. The FF RNA pellet is then rinsed with 70% ethanol and then resuspended in pure water. Wavelengths scans are performed on diluted FF RNA samples (1:10 and 1:20 dilutions in pure water) to assure purity, and the concentration is then collected.

RNA Template + Primer Annealing

A 1:5 ratio of FF-40mer primer:template ($1.8:9.0\ \mu\text{M}$) is created in solution of PBS and MgCl_2 . The sample is then incubated at 90°C for 1 minute and then allowed to cool to room temperature for 30 minutes. The quality of the annealing is measured on a 4–20% gradient TBE gel at 15 W per gel for 30 minutes (all RNA gels are pre-ran at 15 W/gel for 10 minute).

Transcription Assays

$3.6\ \mu\text{M}$ annealed RdRp primer + template dsRNA is placed into 10 separated sample tubes along with RNasIN, $2\ \mu\text{M}$ ATP, and 10x transcription buffer. Varying concentrations of FUS ($0.125 - 2.25\ \mu\text{M}$ as final concentrations, depending on the experiment) are added to each sample, and H_2O and RdRp buffer are added to dilute each sample to a total volume of $20\ \mu\text{L}$ once $3.8\ \mu\text{M}$ RdRp complex solution is added.

Immediately after the RdRp complex is added to the mixture, the samples are incubated at 37°C for 1 hour. The samples are then dyed with FAL-D RNA dye and heated to 90°C for 1 minute. After heating, the samples are loaded to a 15% Urea PAGE gel (that had been previously ran without sample at 15 W/gel for 10 minutes) and the gel ran at 15 W/gel for 30 minutes or until the dye reached the bottom of the gel. The gels were then imaged using Fluorescein.

References

pGBW-m4046802 was a gift from Ginkgo Bioworks (Addgene plasmid # 145683 ; <http://n2t.net/addgene:145683> ; RRID:Addgene_145683

(1) Heydari, S. T., Zarei, L., Sadati, A. K., Moradi, N., Akbari, M., Mehralian, G., and Lankarani, K. B. (2021) The effect of risk communication on preventive and protective Behaviours during the CoVID-19 outbreak: mediating role of risk perception. *BMC Public Health* 21, 1–11.

(2) Programme, U. N. E. (2020) Lessons of Past Coronavirus Pandemics. *Popul. Dev. Rev.* 46, 633–637.

(3) Memish, Z. A., Zumla, A. I., Al-Hakeem, R. F., Al-Rabeeh, A. A., and Stephens, G. M. (2013) Family Cluster of Middle East Respiratory Syndrome Coronavirus Infections. *N. Engl. J. Med.* 368, 2487–2494.

(4) Vannabouathong, C., Davji, T., Ekhtiari,

S., Chang, Y., Phillips, S., Zhu, M., Chagla, Z., Main, C., and Bhandari, M. (2020) Novel Coronavirus CoVID-19. *Bone* 102, 734–44.

(5) Terada, Y., Matsui, N., Noguchi, K., Kuwata, R., Shimoda, H., Soma, T., Mochizuki, M., and Maeda, K. (2014) Emergence of pathogenic coronaviruses in cats by homologous recombination between feline and canine coronaviruses. *PLoS One* 9.

(6) Lau, S. K. P., Poon, R. W. S., Wong, B. H. L., Wang, M., Huang, Y., Xu, H., Guo, R., Li, K. S. M., Gao, K., Chan, K.-H., Zheng, B.-J., Woo, P. C. Y., and Yuen, K.-Y. (2010) Coexistence of Different Genotypes in the Same Bat and Serological Characterization of Rousettus Bat Coronavirus HKU9 Belonging to a Novel Betacoronavirus Subgroup. *J. Virol.* 84, 11385–11394.

(7) Tsoleridis, T., Onianwa, O., Horncastle, E., Dayman, E., Zhu, M., Danjitrong, T., Wachtl, M., Behnke, J. M., Chapman, S.,

Strong, V., Dobbs, P., Ball, J. K., Tarlinton, R. E., and McClure, C. P. (2016) Discovery of novel alphacoronaviruses in European rodents and shrews. *Viruses* 8.

(8) Cotten, M., Lam, T. T., Watson, S. J., Palser, A. L., Petrova, V., Grant, P., Pybus, O. G., Rambaut, A., Guan, Y., Pillay, D., Kellam, P., and Nastouli, E. (2013) Full-genome deep sequencing and phylogenetic analysis of novel human betacoronavirus. *Emerg. Infect. Dis.* 19, 736–742.

(9) Li, T., Liu, D., Yang, Y., Guo, J., Feng, Y., Zhang, X., Cheng, S., and Feng, J. (2020) Phylogenetic supertree reveals detailed evolution of SARS-CoV-2. *Sci. Rep.* 10, 1–9.

(10) Zhang, X. W., Yap, Y. L., and Danchin, A. (2005) Testing the hypothesis of a recombinant origin of the SARS-associated coronavirus. *Arch. Virol.* 150, 1–20.

(11) Lau, S. K. P., Woo, P. C. Y., Yip, C. C. Y., Fan, R. Y. Y., Huang, Y., Wang, M.,

Guo, R., Lam, C. S. F., Tsang, A. K. L., Lai, K. K. Y., Chan, K.-H., Che, X.-Y., Zheng, B.-J., and Yuen, K.-Y. (2012) Isolation and Characterization of a Novel Betacoronavirus Subgroup A Coronavirus, Rabbit Coronavirus HKU14, from Domestic Rabbits. *J. Virol.* 86, 5481–5496.

(12) Yin, W., Mao, C., Luan, X., Shen, D. D., Shen, Q., Su, H., Wang, X., Zhou, F., Zhao, W., Gao, M., Chang, S., Xie, Y. C., Tian, G., Jiang, H. W., Tao, S. C., Shen, J., Jiang, Y., Jiang, H., Xu, Y., Zhang, S., Zhang, Y., and Xu, H. E. (2020) Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* 368, 1499–1504.

(13) Venkataraman, S., Prasad, B. V. L. S., and Selvarajan, R. (2018) RNA dependent RNA polymerases: Insights from structure, function and evolution. *Viruses* 10, 1–23.

(14) 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.

- (15) Schwartz, J. C., Cech, T. R., and Parker, R. R. (2015) Biochemical properties and biological functions of FET proteins. *Annu. Rev. Biochem.* 84, 355–379.
- (16) Schwartz, J. C., Wang, X., Podell, E., and Cech, T. (2013) RNA Seeds Higher Order Assembly of FUS Protein. *Cell Rep.* 5, 918–25.
- (17) Schwartz, J. C., Ebmeier, C. C., Podell, E. R., Heimiller, J., Taatjes, D. J., and Cech, T. R. (2012) FUS binds the CTD of RNA polymerase II and regulates its phosphorylation at Ser2. *Genes Dev.* 26, 2690–2695.
- (18) An, H., Skelt, L., Notaro, A., Highley, J. R., Fox, A. H., Bella, V. La, Buchman, V. L., Buchman, V. L., Shelkovnikova, T. A., Shelkovnikova, T. A., and Shelkovnikova, T. A. (2019) ALS-linked FUS mutations confer loss and gain of function in the nucleus by promoting excessive formation

of dysfunctional paraspeckles. *Acta Neuropathol. Commun.* 7, 1–14.

- (19) Theisen, E. R., Pishas, K. I., Saund, R. S., and Lessnick, S. L. (2016) Therapeutic opportunities in Ewing sarcoma: EWS-FLI inhibition via LSD1 targeting. *Oncotarget* 7.
- (20) Ni, W., Yang, X., Yang, D., Bao, J., Li, R., Xiao, Y., Hou, C., Wang, H., Liu, J., Yang, D., Xu, Y., Cao, Z., and Gao, Z. (2020) Role of angiotensin-converting enzyme 2 (ACE2) in CoVID-19. *Crit. Care* 24, 1–10.
- (21) Shang, J., Ye, G., Shi, K., Wan, Y., Luo, C., Aihara, H., Geng, Q., Auerbach, A., and Li, F. (2020) Structural basis of receptor recognition by SARS-CoV-2. *Nature* 581, 221–224.
- (22) Marfè, G., Tafani, M., Fiorito, F., Pagnini, U., Iovane, G., and de Martino, L. (2011) Involvement of FOXO transcription factors, TRAIL-FasL/Fas, and Sirtuin proteins family in canine coronavirus type

II-induced apoptosis. *PLoS One* 6.

(23) Iserman, C., Roden, C. A., Boerneke, M. A., Sealfon, R. S. G., McLaughlin, G. A., Jungreis, I., Fritch, E. J., Hou, Y. J., Ekena, J., Weidmann, C. A., Theesfeld, C. L., Kellis, M., Troyanskaya, O. G., Baric, R. S., Sheahan, T. P., Weeks, K. M., and Gladfelter, A. S. (2020) Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* 80, 1078-1091.e6.

(24) 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.

(25) Holland, L. A., Kaelin, E. A., Maqsood, R., Estifanos, B., Wu, L. I., Varsani, A., Halden, R. U., Hogue, B. G., Scotch, M., and Lim, E. S. (2020) An 81-Nucleotide Deletion in SARS-CoV-2 ORF7a Identified from Sentinel Surveillance in Arizona

(January to March 2020). *J. Virol.* 94.

(26) Bartolini, B., Rueca, M., Gruber, C. E. M., Messina, F., Carletti, F., Giombini, E., Lalle, E., Bordi, L., Matusali, G., Colavita, F., Castilletti, C., Vairo, F., Ippolito, G., Capobianchi, M. R., and Di Caro, A. (2020) SARS-CoV-2 phylogenetic analysis, Lazio Region, Italy, February–March 2020. *Emerg. Infect. Dis.* 26, 1842–1845.