Analysis Of Respiratory Syncytial Virus Reveals Conserved Rna Secondary Structural Motifs And Impact Viral Lifecycle

Elena M Thornhill, Ryan J Andrews, Zachary Lozier, Elizabeth Carino, Marie-Anne Rameix-Welti, Jean-François Éléouët, Walter N Moss, David Verhoeven

HAL Id: hal-04684010
https://hal.inrae.fr/hal-04684010

Preprint submitted on 2 Sep 2024

1

ANALYSIS OF RESPIRATORY SYNCYTIAL VIRUS REVEALS CONSERVED RNA SECONDARY STRUCTURAL MOTIFS AND IMPACT VIRAL LIFECYCLE

Elena M Thornhill[1,], Ryan J. Andrews[2,], Zachary Lozier[3], Elizabeth Carino[3], Marie-Anne Rameix-Welti[4], Jean-François Eléouët[5], Walter N. Moss[2], and David Verhoeven[1,]*

[1] Department of Veterinary Microbiology and Preventive Medicine, Iowa State University, Ames, Iowa, USA

[2]Roy J. Carver Department of Biophysics, Biochemistry and Molecular Biology, Iowa State University, Ames, IA, USA

[3]Department of Plant Pathology and Microbiology, Iowa State University, Ames, IA, USA.

[4] Service de Microbiologie-Hygiène , Université de Versailles - St Quentin en Yvelines, Hôpital Ambroise Paré, Boulogne-Billancourt, France

[5]Unité de Virologie et Immunologie Moléculaires (VIM), Université Paris-Saclay, INRAE, 78350 Jouy-en-Josas, France

*Corresponding Author
David Verhoeven
College of Veterinary Medicine
Department of Veterinary Microbiology and Preventive Medicine
Iowa State University, Ames, IA, 50010
davidver@iastate.edu
(515) 294-2562

1   **Abstract**

2   An analysis that combined bioinformatics, comparative sequence/structural analysis, and

3   experimental assays has been completed on respiratory syncytial virus (RSV). Both the genomic

4   RNA and its reverse complement were studied using the novel bioinformatics pipeline ScanFold,

5   which predicted 49 regions on RSV RNAs that appear to encode functional secondary structures

6   (based on their unusually stable sequence order). Multiple motifs appear to be conserved

7   between RSV and related virus strains, including one region within the F gene, which had a

8   highly favorable overall prediction metric of a paired secondary structure. This motif was

9   subjected to additional experimental analyses using SHAPE analysis to confirm ScanFold

10   predicted secondary structure. In subsequent analysis, RSV F mRNA transcripts were made by in

11   vitro transcription using T7 polymerase and transcripts which relaxed the predicted secondary

12   structure yielded slightly higher mRNA transcripts and protein expression levels as wildtype F.

13   However, using reverse genetics for comparison of viruses containing wildtype or relaxed F

14   suggested that the predicted secondary structures may be critical for RSV replication in cells. To

15   our knowledge, this is the first study to examine conserved RNA structures across multiple RSV

16   strains and may help identify potential therapeutic targets to inhibit.

17

Introduction

Respiratory Syncytial Virus (RSV) is a significant respiratory viral pathogen that circulates throughout human populations (64 million infections per year) with clinical symptoms manifesting very similarly to influenza virus. RSV typically causes up to 125,000 hospitalizations for bronchiolitis (1,2), 20.3% of the secondary pneumococcal pneumonia cases (3), and is the main viral catalyst for secondary middle ear bacterial infections. RSV is so ubiquitous that almost all children have been infected at least once by two years of age. RSV morbidity is not just restricted to the young. 10,000 deaths are associated with the virus in the elderly per year in the US alone (4). Despite some recent advances in immunogen design, there have been no efficacious vaccine for this virus (5) until recently with two vaccines (RSVpreF3 and RSVpreF from Glaxo Smith and Pfizer had been approved for the elderly (47,48)   . In a study of 72 Gavi-eligible countries RSV caused 1.2 million discounted DALYs, and US$611 million discounted direct costs (6) highlighting RSV's impact on the world economy and population.

The RSV genome is a single-stranded non-segmented negative sense RNA ~15,000 nucleotides in length. The genome encodes 10 genes and 11 different proteins. The 11th protein results from overlapping ORF's in a mechanism that isn't well understood. Transcription proceeds in order along the genome in a start stop mechanism, such that the NS genes should have the highest mRNA transcript levels and with the last gene in the genome L exhibiting the lowest levels of mRNA transcripts (49-50)   . Recently, a new study has shown a non-linear decline in gene transcript aboundance that is not uniform (51)   . Detection primer optimization for each RSV gene, often resulted in the highest CT values for the F gene amplification, indicating a low abundance of transcripts. Given its location in the genome, F should be a mid-abundance transcript. Again, data suggests transcript abundance may be non-linear that is strain

4

1   dependent and thus F may not necessarily be mid-abundance (50)   . Increasing the temperature

2   of the initial RNA reverse transcriptase hexamer binding resulted in in lower CT values, and

3   better detection of the F gene that were more in line with the rest of the RSV genes. These results

4   suggested that some form of RNA secondary structure could be influencing the ability of DNA

5   primers to bind the RNA affecting reverse transcription efficiency. Higher temperatures are

6   capable of relaxing secondary structures, allowing the polymerase better access to the template

7   and to increase reverse transcription efficiency. As such the genomic and mRNA secondary

8   structure of the virus became of interest.

9   Surprisingly, little attention has been focused on the effects of RSV RNA structure

10   (genomic or mRNA) in regulation of either transcription or translation of the virus.

11   Nucleoprotein binds tightly to the genomic and anti-genomic strands making secondary RNA

12   structures in them difficult to form. Thus, here, we concentrate on secondary RNA structure in

13   the mRNA of RSV F transcripts. Multiple mechanisms controlling RSV translation including

14   ribosomal frameshifting and shunting have been proposed (1,9,10). Currently, RNA termination-

15   reinitiation in the M2 mRNA in this overlap region is the most accepted mechanism (11). Work

16   on the M2 gene showed a conserved and essential region that was highly structured. The M2

17   mRNA transcript is translated into the two proteins M2-1 and M2-2 in a not well understood

18   coupled translation mechanism. Due to the placement of the conserved structure, it is likely that

19   the secondary RNA structure is causing the coupled translation event. This structure indicates

20   that there could be other such RNA secondary structures within other mRNA transcripts, the

21   genome, or the anti-genome that influence the regulation of gene expression in RSV (10).

22   While the work done on M2 is promising, there are precious few other papers looking

23   into other possible sites of secondary RNA structures such as within F that might be influencing

1   gene regulation in RSV. In other RNA viruses, like Dengue and Influenza A, a number of host

2   binding proteins regulate viral replication by enhancing, blocking, or helping with RNA

3   movement around the cell (12-14). However, their genomes are different from mononegavirales

4   and not as highly encapsidated by the nucleoproteins. Previous analyses of influenza A virus

5   (IAV; another negative-sense RNA virus) have discovered numerous structural elements that are

6   implicated in critical processes (15). In the IAV genomic RNA long-range base pairing

7   interactions form a "panhandle" secondary structure that is essential for forming the promoter for

8   replication (16). In IAV coding RNAs, motifs have been discovered that regulate alternative

9   splicing and that are critical to viral fitness (17) but IAV is different in that it replicates in the

10   genome and thus any structures in RSV's transcripts would not likely be for splicing. Other

11   motifs have apparent functions in the regulation of translation (18) or that may be able to affect

12   genome packaging (19).

13   While it is suggested that RSV's M (20) and known that RSV's M2-1, L, and N (genomic

14   RNA only) proteins (21,22) are all self-viral RNA binding proteins (either directly or indirectly

15   through interaction with another protein), not much is known about RSV's RNA structure and

16   host protein (if any) binding to genomic or mRNA. RNA folding is achieved by intra-molecular

17   interactions between the nucleotides and are often sites of specific interactions with proteins. A

18   similar approach to the one used recently for analysis of ZIKA and HIV genomes using

19   ScanFold-Fold (23) was used across a multitude of RSV viral strains (A and B) to identify key

20   sites of structural conservation. These sites may identify key sites of interactions with host or

21   viral proteins critical to viral replication that could be manipulated to prevent or limit RSV

22   infection. Again, caution with this statement should be taken as the genome is bound by the

23   nucleoproteins.

1   In this study, we found a conserved viral RNA structure in the mRNA of the F

2   gene through experimental and computational methods. This structure, which also is indicated

3   computationally to have a mirror in the genomic RNA, appears to be critical in the RSV lifecycle

4   as relaxation prevented viral rescue. Further understanding of RNA structures' roles in regulating

5   viral replication or protein or gene expression may foster therapeutic targets for this significant

6   human pathogen.

7

8

1   **Materials and Methods**

2   **Viral Strains:** RSV A (A2, 2001, 2006, 1997, 1998, Memphis-37) and RSV B (B1)

3 strains were obtained from Beiresources (Mananas VA) and expanded 1 time in HEp2 cells

4 (ATCC, Mananas, VA). Virus was purified according to (24) with the following modifications.

5 NT buffer was used to make 50% PEG 8000. 10% of the PEG solution was added to the viral

6 supernatant and incubated with agitation at 4°C for 1hr followed by centrifugation at 5000xg for

7 40min. The resulting pellet was re-suspended in DMEM with 3% sucrose.

8 **Plasmid Construction:**

9   -   **PCR2.1 containing the F gene-** Constructed using a TA cloning (Invitrogen,

10   Carlsbad, CA, K204001) system according to manufacturer's specifications and a

11   PCR amplified F gene containing fragment.

12   -   **Partial F Relaxed**- The relaxed F gene was generated by site mutagenesis of PjetF

13   (25) in order to relax the paired hairpin within the middle of the F gene but retain a

14   similar codon optimization as the original F. Specifically, the sequence

15   ATTAAGCAAGAAAAGGAAAAGAAGATTTCTTGGTTTT at base pair 385 of F

16   (within PjetF) was mutated to

17   CCTGAGTAAAAAGCGCAAGCGCCGGTTCCTCGGCTTCTTC thus retaining the

18   amino acid sequence TLSKKRKRRFLGFL.

19   -   **PjetF-relaxed**- Constructed from the RSV PjetF described in (25) and a synthesized

20   fragment of F containing the relaxed sequence. The mutated F fragment was cloned

21   into the parent PjetF via HiBuilder (NEB, Ipswich, MA) and was sequenced.

22   -   **pACNR–rHRSV-relaxed**- Constructed using the newly constructed PjetF-relaxed

23   and the pACNR–rHRSV described in (25) via Hibuilder (NEB, Ipswich, MA) and

24   sequenced.

1 **Scanning window analysis and structural motif discovery:**

2   The negative-sense RSV genomic RNA (NCBI: JQ901447.1), and it's positive-sense

3 coding RNA counterpart were both analyzed using ScanFold; a pipeline which consists of a

4 scanning window analysis which is subsequently analyzed to highlight secondary structures

5 which are most likely to be functional (26). The scanning window analysis (ScanFold-Scan) was

6 run using a 120 nt window size, 1 nt step size, and several folding metrics were recorded

7 (described in detail in (26). Two key metrics are the minimum free energy (MFE) and its

8 associated z-score. The MFE was predicted with RNAfold (version 2.4.3) as it corresponds to the

9 RNA secondary structure which would yield the lowest free energy during its formation

10 (according to the Turner energy parameters (26).The corresponding thermodynamic z-score

11 reports how unusual the MFE is for its particular nucleotide composition: by comparing the

12 native MFE (MFEnative) to one hundred randomized sequences with the same mononucleotide

13 content (MFErandom) and normalizing by the standard deviation of all MFEs (σ), thus indicating

14 how many standard deviations more stable (i.e. negative z-scores) or less stable (i.e. positive z-

15 scores) the MFEnative is than expected. Subsequently, ScanFold-Fold (26) was used to identify

16 base pairs which consistently contributed to the formation of unusually stable RNA secondary

17 structures. This is accomplished by generating a z-score weighted consensus secondary structure

18 from all base pairs predicted across all analysis windows.

19 **Comparative sequence and secondary structure alignment:**

20   A total of 1222 Human RSV genomes were downloaded from NCBI Nucleotide database

21 following a search for whole genomes for "Human orthopneumovirus". Genomes were aligned

22 to the ScanFold analyzed genome (accession JQ901447.1) with default settings using the

23 MAFFT web server (27,28). The list of ScanFold-Fold identified base pairs (with Zavg < -2 from

1   JQ901447.1) were compared to the alignment to determine the extent of structural conservation

2   for each base pair throughout the alignment.

3   **qRT--PCR of Viral Genes:**

4   Virus RNA was extracted by Viral RNA/DNA isolation kit (Invitrogen, Carlsbad, CA)

5   according to the manufacturer's instructions. Primers for each gene were created on Primer-Blast

6   (NCBI, Bethesda MD) and synthesized (IDT, Iowa City, IA). The sequences are as follows,

7   Highly structured region (Primer1) F: gcaaagctgcagcatatcaa R: cattgttggacattaacttttctg

8   Unstructured region (Primer 2) F: agaagtcttagcatatgtggtac R: aacagatgttggacccttcc RNA diluted

9   at 1/100 was amplified using Luna® Universal One-Step RT-qPCR Kit (NEB, Ipswich, MA)

10   according to the manufacturer's directions on a QuantStudio3 (Applied Biosystems, Foster City,

11   CA). These primers bind around the B structure in Figure 1.

12   **SHAPE Analysis of Predicted Structure:**

13   RNA was generated through InVitro Transcription according to (Invitrogen, Carlsbad,

14   CA). 3μg of RNA was used for each SHAPE sample. After modification of the RNA was

15   complete samples were reverse transcribed into cDNA using Superscript 3 according to

16   manufactures specs (Invitrogen, Carlsbad, CA), with 1ul of a 1/10 of a 100uM stock reverse

17   fluorophore tagged primer instead of the random hex/oligo dt. The sequencing ladder sample

18   used a ddntpA mix in place of the dntp mix for use as a ladder. Protocol from here is as

19   described in (29) re-suspending the final pellet in 10ul. Reverse primer sequence is as follows

20   Set 1: cagccttgtttgtggatagtagag tagged with FAM. Samples were sent to be sequenced by

21   capillary electrophoresis and then analyzed using QuSHAPE and RNAFold (30,31).

10

Generation of SHAPE Figure:

Done by mapping reactivity values onto the RNAFold structure generated from the SHAPE reactivity values or the ScanFold generated structure based off the models in (26) via RNA2Drawer (32).

PCR of SHAPE Products:

2ul of SHAPE produced cDNA and of cDNA produced from an unmodified FmRNA and cellular mRNA using reverse primers designed to target structured and unstructured regions were PCR'd according to manufacturer's specifications (Takera, Mountain View, CA). PCR products were then run on an agarose gel. Primer sets are as follows, for cDNA made with the primer for the structured region Primer3 F: gtgttggatctgcaatcgcc R: gacttcctcctttattattgacatga with the unstructured region primers Primer4 F: agaagtcttagcatatgtggtac R: tttattggatgctgtacatttggt. Primer3 created cDNA was run at 60°C as was Primer4 as per required TM for primer sets for 20 cycles.

Cell Transcription and Transfections:

PCR products were generated using the PjetF and PjetF-relaxed as templates and the following primer set which contains a T7 promotor sequence Forward: aaataatacgactcactatagggccatggagttgccaatcctcaaaacaaatgc Reverse: tcagctactaaatgcaatattatttataccactcagttgatcc. PCR products wildtype and relaxed were then transfected into T7 expressing Baby Hamster Kidney Cells (33) (BHK: BSRT7) using the X2 transfection system according to manufacturer specifications (Mirus, Madison, WI). Transfections were incubated overnight, and cells were harvested. mRNA was extracted using RNeasy Plus according to manufacturer specifications (Qiagen, Hilden, Germany). Extracted RNA was treated with DNaseI to remove any residual contaminating DNA according to manufactures specifications (Thermo Fisher Scientific, Waltham, MA). Samples and standards

11

1 were then amplified using the Luna 4x UDG system (NEB, Ipswich, MA) and the following

2 primer set (Forward: tgcagtgcagttagcaaagg Probe: FAM/cagaattgc/ZEN/agttgctcatgca/ 3IABκFQ

3 Reverse: gattgttggctgcttgtgtg) according to the manufacturer's directions on a QuantStudio3

4 (Applied Biosystems, Foster City, CA). Copies per ul were calculated using Linear interpolation

5 and the cts generated by the standards.

In vitro transcriptions and RNA pulldowns

7 Plasmids containing either the RSV F in wildtype sequence or relaxed were used to

8 generate RNA using biotin labeled UTP at 1/3 the amount of unlabeled UTP in a Biosearch

9 Technologies (Middlesex UK) transcription kit. To minimize contributions from other structures

10 or the UTRs, we only generated short transcripts (still predicted to fold correctly) around our

11 predicted structure in region B of figure 1. RNA was purified by RNA purificiation kit (NEB,

12 Boston MA). RNA was tranfected in Hep2 cells using RiboJuice (Millipore, Temecula CA)

13 according to their instructions. Cells were then fixed with UV using a stratalinker and cells

14 disrupted by sonication in the presence of protease inhibitor (HALT, Thermofisher, Waltham

15 MA). Anti-bioin magnetic beads were reacted and RNA pulled down by magnet. Pull-downs

16 were added to PAGE gel loading dye and boiled before loading on PAGE gels (Invitrogen,

17 Carlsbad CA), transferred, and blotted using antibodies (Genetex, Irvine CA) to identified

18 binding proteins.

Imaging of PCR Transfections:

20 Duplicate samples of those harvested for RNA were harvested and fixed in solution. Cells

21 were stained for F using MPE8 (5) and an anti-human Alexa Fluor 555 secondary. Imagining

22 was done by dotting a small amount of the stained cell solution on a slide and using a fluorescent

23 scope.

1 **Construction and Rescue of RSV using Reverse Genetics:**

2 This was done according to prior published studies (25) with the following modifications:

3 Initial transfections were done using the X2 system and were scaled up 2.6x and dropped on to

4 slightly under-confluent BSRT7 cells in a 25cm2 flask. Both pACNR–rHRSV-relaxed and

5 pACNR–rHRSV transfections were left on cells for 4hrs before cells were washed twice and new

6 media was added to the transfected cells. Transfected cells were incubated for ~4 days or until

7 media color indicated it needed to be changed or cells became 100% confluent. Both virus and

8 infected cells were harvested and moved to a larger flask after this time. Virus was sucrose

9 purified after a final incubation in a 185cm2 flask and 200ul were taken for viral RNA extraction

10 and to be plated onto new HEp-2 cells for staining.

11 **Intracellular Staining of wildtype and relaxed Virus:**

12 Viral infection was incubated for 24hrs before cells were washed and fixed in

13 paraformaldehyde. Cells were washed in 1x Permwash twice and incubated in Permwash for 30

14 min. at 4°. Cells were stained for RSV using a human anti-RSV antibody (Beiresources,

15 Mannasas VA) and an Alexa Fluro 555 secondary (Invitrogen), were re-fixed and imaged.

16 **Surface Staining of wildtype and relaxed Virus:**

17 Infected BSRT7 were harvested at the same time as viral harvest and were surface stained

18 for F using the MPE8 (51) antibody and anti-human Alexa Fluor 555 secondary . Imagining was

19 done by dotting a small amount of the stained cell solution on a slide and using a fluorescent

20 scope.

21 **Copies per ul for wildtype and relaxed Virus:**

22 Viral RNA was extracted using the Nucleospin system according to manufactures

23 specifications (Macherey-Nagel, Düren,). Extracted RNA samples and DNA standards were then

24 amplified using One Step PrimeScript™ RT-PCR Kit (TAKARA BIO INC, Kusatsu, Shiga,

13

1   Japan) and the following primer set ((Forward: atggacacgacacacaatga Probe:

2   FAM/ccacaccac/ZEN/aagactgatgatcaca/ 3IABκFQ Reverse: gtggcctgttttcatcaag) according to the

3   manufacturer's directions on a QuantStudio3 (Applied Biosystems, Foster City, CA). Copies per

4   ul were calculated using linear interpolation of the CTs generated onto a standard curve of a

5   plasmid containing the NS1 gene.

6   **Statistics:**

7   Student T tests were run on the F gene expression of plasmids and viral infection for

8   predicted structure and unstructured regions. Student T tests were run on the mRNA

9   transfections between log copies per ul of wildtype vs relaxed mutants. Student T tests were run

10   on the log copies per ul of viral concentration recovered from wildtype vs relaxed mutants. A p-

11   value of less than 0.05 was considered significant.

12

13

<div align="center">

Results

</div>

Scanning window analysis

The RSV genome was analyzed using the ScanFold pipeline with a window size of 120 nt and step size of 1 nt. This resulted in 15088 overlapping windows spanning both genomic and complementary (anti-genome/mRNA) RNAs (Figure 1). For each window several RNA folding metrics were predicted, which can be used to assess the sequence's capacity to form an RNA secondary structure. Two metrics of particular interest are the predicted minimum free energy (MFE) of the window sequence and its associated z-score. The resulting MFE and z-score values for each window are plotted vs. the RSV genome in Figure 2. Across the coding sequence, native MFE values range from -36.7 to 0.0 kcal/mol, averaging -15.67 kcal/mol; across the genomic sequence native MFE values range from -40.4 to -3.4 kcal/mol, averaging -18.98 kcal/mol. Across the genome, thermodynamic z-scores ranged from -4.68 to 2.57 and -5.19 to 2.50 in the coding and genomic strands respectively. Despite having a more positive max and min z-score, the average thermodynamic z-score of the sequence was slightly more negative (-0.44) than the genomic sequence (-0.36) suggesting that overall, the coding sequence has more bias for being ordered (by evolution) to form stable secondary structures—indicating potential function. This was also observed in IAV (35) where regulatory elements were preferentially found in the coding RNA vs. the genomic RNA, which does not undergo complex regulation.

RNA structural motif discovery

In order to determine the specific structures that could be giving rise to low z-score regions, the ScanFold-Fold algorithm (26) was employed. This algorithm systematically analyzes the MFE structure and z-score from each window and identifies base pairs which were consistently present in low z-score windows. In the strand, 29 RNA secondary structures were predicted to contain at least one base pair which consistently appeared in windows averaging z-

1   scores less than -2 (Zavg< -2); with only 13 structures (greater than 4 bps in length) composed

2   entirely of Zavg<-2 base pairs (Figure 1). These z-score generating structures comprise five

3   structured regions which are found within four RSV mRNAs (here, structures are considered part

4   of the same region if they are within 120 nts of each other). The first structured region consists of

5   a single hairpin which spans the coding and 3' untranslated region (UTR) of the SH mRNA. The

6   next three structured regions occur throughout the coding region of the F protein mRNA. The

7   fifth structured region is relatively large (205 nt long), spanning three different mRNA regions,

8   and consisting of 7 hairpins composed entirely of Zavg<-2 bps and (ranging from 14 bps to 3 bps

9   in length). This region is remarkable for its overlapping genetic roles; whereby it comprises the

10   3'UTR of M2-1, the stop codon for M2-2, and the 5'UTR and start codon for the L protein. This

11   region also had the highest level of base pair conservation at 97.2% (as opposed to an overall

12   average of 90.0% for all Zavg<-2 base pairs).

13   In the genomic strand of negative polarity, 20 RNA secondary structures were predicted

14   to contain at least one bp with Zavg< -2. Of these, 14 were composed entirely of Zavg< -2 bps.

15   comprising seven structured regions throughout the genome. The majority of these structured

16   regions were located in different locations when comparing the genome locations and the coding

17   sequence (Fig. 1 F, G, H, J, and L) with only two regions mapping to structured region locations

18   in the coding strand (Fig. 1 I and K).

19   Since we initially were having different CT values detected when trying to amplify off

20   the same cDNA strands for the RSV F mRNA, we decided to look more indepth whether

21   conserved structures in this region might be impeding our amplifications of this transcript. As

22   shown in Figure 2, the RSV F mRNA region (and mentioned above), had three conserved

23   regions across all strains of RSVa.

Template dependent amplification shift indicative of RNA Structure

Using the data obtained above, primers were designed to target the regions of 6273-6478 (Primer1) and 6528-6647 (Primer2) of the F transcript. The region targeted by Primer1 being predicted to be highly structured and Primer2 region being to be lightly structured or unstructured. The primers were used to generate cDNA from HEp-2 cells infected with RSV, uninfected Hep2 cells, and plasmid DNA from a PCR2 plasmid containing the F gene. The primer binding site had an effect on the degree of PCR amplification (Fig 3). Infected Hep2's showed a difference in amplification with the Primer1 located in a region having higher CT's than that of the Primer2 region. Higher ct's correspond to a lower amplification/gene expression, while a lower ct correlates to a higher amplification/expression. Differences in the amplification of the two regions were significant as seen in the difference in relative fold increase over universal uninfected cells. Amplification on our RSV F plasmid showed no significant difference in the difference in relative fold increase over universal uninfected cells between the two regions. The plasmid DNA is not only double stranded and thus not structured, but was never subject to the inefficiencies in the reverse transcriptase that the viral cDNA was. Because of this, the plasmids should not have a significant difference between the two primers as there is no difference in the difficulty in amplification between the two sections that was predicted to occur in the viral RNA conversion to cDNA. This shows that in regions of predicted as structured there is a significant difference in amplification, not attributed to primer efficiency, giving more credence to the presence of a structured region in the F gene that may affect regulation of the gene.

Experimental confirmation of computationally predicted RNA structure in F mRNA

Modified and Unmodified cDNA are differently amplified

17

1   PCR products created from SHAPE generated cDNA confirmed that the cDNA created was

2   indeed F and that the amount of full-length transcripts that were produced off of the modified

3   RNA in the structured region was lower than that of the unmodified RNA of the same region as

4   seen in Figure 3 This difference was not seen in the modified and unmodified RNA generated

5   from the unstructured region. Variable amplification of cDNA through the structured region

6   showed that the modified RNA had the lowest number of transcripts that ran through the

7   structured region and that the ladder cDNA which was the same concentration as the cDNA from

8   the modified RNA had more full-length transcripts as seen by the darker band. This was not the

9   case for the cDNA created from the non-structured region where all of the created cDNA has

10 similar amplification. This further gives evidence to the RNA being structured at the Primer3

11 location as modification is intended to create truncated cDNA products where structure exists on

12 the RNA, and non-structured regions are supposed to be able to produce full length products.

13 Stable RNA structure is also known to cause truncation issues due to the enzyme falling off of

14 the template.

15   ***Structure determination:*** Experimental determination of F mRNA secondary structure

16 was determined through SHAPE probing using benzoyl cyanide modified RNA. The primer set

17 used was designed to target the region of 6036-6134nt of the F transcript that the ScanFold

18 results predicted a stable and very conserved structure. Normalized SHAPE reactivity values

19 close to zero indicate a low reactivity and a high propensity to form base pairs whereas high

20 reactivity values indicate highly modified bases that are likely to be unpaired because the

21 benzoyl cyanide is able to access the nucleotides for modification. Areas of double stranded

22 RNA inhibit modification and thus have lower reactivity. For our purposes reactivity cutoffs

23 were defined as follows 0-.4, .4-.6, >.6 and were taken to be unreactive (base-paired), moderately

1   reactive (intermediate levels of base pairing), and highly reactive (unpaired) respectively (36-

2   40). SHAPE analysis was performed four times and the final normalized reactivity values

3   generated from QuSHAPE of all sets were averaged and run in RNAFold to generate a predicted

4   structure using the generated SHAPE restraints. Reactivity values were obtained for more than

5   90% of nucleotides in the target region sequence. Said values were then mapped onto the

6   RNAFold and ScanFold generated structures in RNA2Drawer (32) to determine the placement of

7   the normalized reactivity values. While the exact base pairing predicted by each model is not

8   identical both produce two hairpins directly adjacent to one another in the predicted region.

9   Additionally, within the second hairpin both models show identical tips and a large bulged out

10   region that are highly modified (Figure 4). The four individual SHAPE data sets were also

11   independently run on RNAFold and 75% of them generated the bulge and 100% had the same tip

12   (data not shown).   The regions that differ on the first hair pin contain a stretch of U's that may

13   "breath" and allow the real structure to alternate between a more linear and double stranded

14   orientation and a bulged out of single stranded conformation. RNA structures in their

15   biologically relevant environments are rarely static and often change conformation (41,42).

16   Overall, the experimental data supports the existence of two hairpin structures in the region

17   predicted by the computational model.

18   **Production of F Protein Unaffected by RNA Sequence Change:**

19   First, we generated a relaxed version of the F gene on our plasmid to generate a change in

20   mRNA structure at the same region as Fig 3 around the structure in region B of figure 1 using a

21   shot-gun approach making a major change to the RSV structure located in this region. The amino

22   acids remained the same and we tried to keep the codon usuage roughly the same.

23   To confirm that the modifications made to the RNA sequence of F and the subsequent

24   relaxation of the secondary RNA structure in the mRNA did not affect the production and

1 transcription of the F protein/mRNA we analyzed the levels of mRNA and protein produced

2 from T7 BHK cells transfected with T7 promotor containing PCR products of either the wildtype

3 F or the relaxed F (Fig 5A-C). The mRNA quantity was analyzed by qRT-PCR and indicated

4 slightly higher levels of Relaxed mRNA were produced with slightly higher amounts of the

5 relaxed version (Relaxed) being produced (Fig 6A). The copy number of produced mRNA

6 correlated with the protein production of F as visualized by intracellular staining slightly higher

7 in the relaxed version (Fig 6B).

8 **RNA Structure relaxation prevents viral rescue:**

9 To further test whether altering the secondary structure would have an effect on viral

10 production given that the T7 polymerase system is very robust at generating mRNA off DNA

11 with the correct promoter, we used the same wildtype, relaxed region, and a region with a

12 compensatory mutation in our shuttle vector that we then moved into a plasmid containing the

13 viral genome (Fig 7A). Initially produced viral titer is very low and often below detection. As

14 such we not only passaged the virus containing supernatant but the initially transfected cells so

15 that any infected cells had the optimal opportunity to produce virus and to allow infection to

16 spread to neighboring cells. Redistribution of the cell monolayer by splitting cells into a new

17 flask also allows the viral infection to better spread through the flask as infection clusters (cells

18 that are infected and their immediate neighbors) are redistributed so that new infection centers

19 can form, and the virus may spread to new uninfected cells. As such, the relaxed RSV virus was

20 given all available opportunities to rescue both in time of incubation and repeated passaging in

21 multiple flask sizes. While the wildtype virus consistently cased syncytia and spread, the relaxed

22 failed to do so. After ~ 6 days the wildtype infected cells blew out causing a lifting of the

23 monolayers and large quantities of syncytia cells were seen floating in the media. In contrast the

24 relaxed infected cell monolayer as intact and the cells appeared only slightly stressed as did the

1. negative control cells likely due to the pH of the media (data not shown). Dectection of virus

2. after transfer of viral supernatant on uninfected cells after transfection of the reverse genetics

3. system, was done by intracellular antibody staining. We could see high levels of RSV within

4. cells in our compensatory mutant and wildtype viruses but much lower levels of F staining in our

5. relaxed F rescue (Fig 7B). We then took the cell supernantants to determine how much virus was

6. being released between the wildtype and relaxed and found a significant difference in virus with

7. relaxed failing to replicate and release (Fig 7C). These data were very interesting as we orgnially

8. thought from Fig 6 that relaxed would have lead to higher viral production.

9. **Protein binding prediction identified a number of potential partner proteins interacting**

10. **with this region of the F mRNA.**

11. We fed the structured region into RBPsuite and found a number of predicted proteins that

12. could bind to this region of the RSV F mRNA (Figure 8A). We next generated RNA off our

13. wildtype and relaxed F plasmids using in vitro transcription. We used biotin labeled UTP to

14. allow us to capture the mRNA generated after we transfected the mRNA into Hep2 cells. After

15. ruling out a number of predicted proteins (Pum2, MBL1 etc, not an exhaustive search), we

16. probed on a western blot for EIF4B, we found a positive signal in the wildtype but not the

17. relaxed site after magnetic pulldown on the RNA (Fig 8B). However, there could be other

18. proteins binding that are subject to future studies. We only used a short version of the F

19. transcript centered around the predicted folding and not be influenced by the parts of the mRNA

20. (structured or not). Thus, EIF4B, which is expected to be on mRNA, would not have shown up

21. in our experiment unless it was binding to our structure.

22.

Discussion

RNA secondary structure often serves an important regulatory purpose in mRNA transcripts, such as ensuring transcript stability and transport or influencing the timing or level of expression. Some secondary structures like those in the M2 mRNA of RSV are thought to allow access to alternative reading frames and allow for additional proteins to be produced through mechanisms such as ribosomal frameshifts. Other viruses use secondary structure to recruit RNA binding proteins which are crucial to their replicative life cycle (43,44).

ScanFold predicted the presence of conserved stable structures in nucleotides 6036-6134, 6366-6401, and 7230-7317 of the F transcript. This region corresponds to the coding region of the F gene, and is present in the F mRNA transcript. This region also corresponds to the fusogenic domain of F possibly explaining why this region may have more conservation than others. While the fusogenic domain region is important for the F protein and subsequently the conserved structure is on the RNA level. RSV has a high level of RNA sequence variation with the majority occurring in the G gene (45,46) but while sequence variability can affect protein viability, RNA secondary structure is not strictly dependent on the RNA sequence. Conservation in structure across viral strains often indicates a function within the viral lifecycle. The predicted structure information was used to design experimental tests.

qPCR data generated from primers designed to target predicted areas of structured and unstructured regions indicated the presence of a secondary RNA structure that inhibited the action of the reverse transcriptase resulting in a higher ct value than in areas of no structure. The higher ct of these regions was not found in the qPCR of plasmids containing the F gene indicating the difference was not due to primer inefficiency.

While the PCR amplification of the SHAPE generated RNA of the non-structured region was uniform regardless of whether the RNA was subjected to modification by benzoyl cyanide

1   or not, the amplification of the SHAPE generated RNA from the structured region was not.

2   cDNA off of the modified RNA amplified fewer full-length transcripts than that of the

3   unmodified RNA or especially that of the ladder. The uneven distribution of full-length transcript

4   amplification is further proof of the secondary structure existing within that region and that the

5   SHAPE generated structure is indeed from F.

6   The structure found conserved in multiple RSV strains through ScanFold predictions was

7   also found through experimental SHAPE probing giving high credibility to the presence and

8   conservation of the structure. Highly conserved structures are often essential and thus can be

9   targets for viral therapeutics. The F protein of RSV is an essential protein, as it allows for cell

10   entry and is expressed on the viral envelope.

11   When this structured region of the RNA sequence was mutated to relax the secondary

12   structure PCR transfection showed that the mutation did not alter the production of mRNA nor

13   that of the F protein. This coupled with the failure to rescue the relaxed version of the reverse

14   genetics RSV virus indicates that there is likely some regulatory function in the F mRNA

15   lifecycle that the secondary structure serves as on its own the mutated sequence is able to

16   produce F. While the mutation likely affects the mRNA, the computational analysis of both the

17   coding and genomic sense of the F region indicates an almost twin structure located in the same

18   region of F on the genomic strand. Mutation of the region to relax the mRNA structure would

19   have also mutated the genomic structure. As such we can only say for sure that the relaxation of

20   this structure in either/both the mRNA or/and the genomic region of F lead to the failure to

21   rescue the relaxed virus.

22   The exact purpose of the secondary structure will be elucidated in further work but it is

23   tempting to speculate what that purpose may be. Given that both the wildtype and relaxed

23

1   mRNA are capable of producing the protein but the relaxed virus fails to rescue it is likely that

2   the mutation causes a defect in a yet unknown regulatory mechanism that controls F expression,

3   transcription, or trafficking within the context of a native infection. The RSV M2-1 protein may

4   play a role as it is known to bind to nascent viral mRNAs not just at the polyA tail but also

5   across the length of the transcript with a preference for A/U rich sites. Having bound the nascent

6   mRNA, M2-1 can either leave the inclusion body and exit into the cytoplasm where it releases

7   the mRNA for translation or go into a sub structure of the IB called an IBAG (52)   . The

8   composition of viral RNAs and the duration of their stay are located in the IBAG, and as such it

9   is a possible site of viral regulation. Perhaps the structure contributes in some way to a

10 sequestration of the F mRNA in the IBAG. We are also unsure if there are more RNA binding

11 proteins interacting with the structured regions in RSV F mRNA and do not quite know why

12 EIF4B would be binding there since this translational helper usually binds IRESes or the

13 beginning of transcripts. It is speculative but we would suggest that perhaps EIF4B might help

14 the translational machinery stay on the transcript but given that relaxed T7 polymerase produced

15 F had more expression this remains to be explored. Further work on exploring if EIF4B is

16 needed or if other factors not yet identified but predicted are important to translation of F protein.

17   One caution we are aware of with our relaxed F structure work is the shift from adenine

18 to guanine or cytosines. These subsitions could affect internal methylation of the mRNA and

19 thus its stability. Further studies to slowly walk nucleotide point mutations are underway.

20   While we have examined potential RNA folding the RSV genome, nucleoproteins are

21 known to tightly associate with the RNA and might prevent these RNA stuctures from forming.

22 However, secondary RNA structures are known to occur in influenza virus, albeit it's a

23 segmented virus for which RSV is not, and nucleoproteins are tightly associated with that

1   genome as well (53-54)   . However, nucleoprotein binding the the influenza genome is not

2   uniform and a similar thing may occur in RSV (55)   . Thus, much more exploration will need

3   to be focused on validating these structures in the RSV genome and their role if found.

4   Secondary RNA structures in mRNA as described here would not be interfered by

5   nucleoproteins and are well known to occur. This is the first paper that we are aware of that

6   assigns highly conserved RNA structures in the RSV F transcript and a role in pathogenesis.

7   Many more structures are left to be explored.

8

1

Figure Legend

Figure 1: Scanning window/RNA structural motif discovery and analysis of the RSV genome and antigenome/mRNA transcripts. 1222 sequences of RSV were fed into ScanFold and areas of predicted (z-score >3) were plotted. Antigenome or mRNA transcripts are shown above the genome predicted structures.

Figure 2: Scanning window/RNA structural motif discovery and analysis in the F mRNA region. A) ScanFold generated MFE and z-score values plotted against the RSV mRNA. B) Conserved structure predictions for the F mRNA similar to figure 1.

Figure 3. **Primer binding site has an effect on the degree of PCR amplification.** Primer1 corresponds to a predicted highly structured region and the Primer2 corresponds to a predicted low/unstructured region.you should described how were done the experiments and analyzed the results. How were quantified the PCR products? Is it Q-PCR or gel analysis?

Figure 4: Experimental Determination of Secondary RNA Structure. A) PCRs of F off of SHAPE modified and unmodified cDNA using primers for structured and unstructured regions. Primer3 set is a region of predicted structure and Primer4 set is predicted to be unstructured. Modification is intended to create truncated cDNA products where structure exists on the RNA, and non-structured regions are supposed to be able to produce full length products. **B)** Normalized SHAPE reactivity values overlaid on the lowest MFE structure generated by RNAFold using the SHAPE reactivity restraints. 0-.4, .4-.6, >.6 were taken to be unreactive (base-paired), moderately reactive (intermediate levels of base pairing), and highly reactive (unpaired) respectively. **C)** Normalized SHAPE reactivity values overlaid on the computationally predicted structure shown in Figure 2 region M1.

1 **Figure 5: Design of relaxed mutation for F.** We concentrated on the same M1 region form

2 Figure 2 and determined the amino acid sequence. **A)** We then designed a relaxed mutant

3 keeping the same amino acids and a similar codon usuage (Genscript table) **B)** but only focusing

4 on ½ of the conserved structure (encircled in black). **C)** RNAfold prediction showed that the

5 relaxed region should destroy the secondary RNA structure in RSV F mRNA. **D)** Interestingly,

6 this region aligns with the highly conserved furin cleavage site in RSV F protein. WT: wildtype,

7 Relx: relaxed mutant.

8 **Figure 6: F Protein and mRNA of WT and Relx F. A)** Calculated copies per µl on a log10

9 scale of mRNA harvested off of either PCR fragment wildtype or relaxed transfected T7BHK

10 cells. **B)** Anti-RSV stained cells transfected with either wildtype or relaxed F PCR fragments.

11 Red is detected virus.

12 **Figure 7: Relaxed RSV F mRNA impairs viral pathogenesis after viral rescue.** We used our

13 wildtype virus, our relaxed virus, but also generated a compensatory mutant in the same RSV

14 region to restore the RNA structure to the same shape. **A)** RNAfold predicted structures are

15 shown for comparison. **B)** We then used our reverse genetics system to rescue virus after

16 transfection and transfer virus to new uninfected Hep2 cells. These were stained intracellularly

17 with anti-RSV antibody and a fluorescent reporter. **C)** Viral supernatants were harvested and

18 subjected to qRT-PCR to determine viral copy numbers. Neg: uninfected cells, other

19 abbreviations similar as prior figures.

20 **Figure 8: Predicted RNA binding proteins within the structured M1 region of RSV F**

21 **mRNA. A)** We used RBPsuite to feed in the sequence as outlined in the figure and generated a

22 table of predicted proteins that could bind to these structured regions. **B)** RNA pulldowns after

23 transfection of sham (unlabeled wildtype), wildtype, or relaxed F RNA suggest at least that

27

1   EIF4B might bind to the region. Pulldowns were repeated 6 subsequent times and 2

2   representatives are shown in 1 blot.

3

4

5

6

7

8

28

1   **References**

2   1.   Collins, P.L., Fearns, R. and Graham, B.S. (2013) Respiratory syncytial virus: virology,
3   reverse genetics, and pathogenesis of disease. Curr Top Microbiol Immunol, 372, 3-38.

4   2.   Piedimonte, G. and Perez, M.K. (2014) Respiratory syncytial virus infection and
5   bronchiolitis. Pediatr Rev, 35, 519-530.

6   3.   Weinberger, D.M., Klugman, K.P., Steiner, C.A., Simonsen, L. and Viboud, C. (2015)
7   Association between respiratory syncytial virus activity and pneumococcal disease in
8   infants: a time series analysis of US hospitalization data. PLoS Med, 12, e1001776.

9   4.   Falsey, A.R. and Walsh, E.E. (2005) Respiratory syncytial virus infection in elderly
10   adults. Drugs Aging, 22, 577-587.

11   5.   Thornhill, E.M., Salpor, J. and Verhoeven, D. (2020) Respiratory syntycial virus: Current
12   treatment strategies and vaccine approaches. Antivir Chem Chemother, 28,
13   2040206620947303.

14   6.   Li, X., Willem, L., Antillon, M., Bilcke, J., Jit, M. and Beutels, P. (2020) Health and
15   economic burden of respiratory syncytial virus (RSV) disease and the cost-effectiveness
16   of potential interventions against RSV among children under 5 years in 72 Gavi-eligible
17   countries. BMC Med, 18, 82.

18   7.   Fearns, R., Collins, P.L. and Peeples, M.E. (2000) Functional analysis of the genomic and
19   antigenomic promoters of human respiratory syncytial virus. J Virol, 74, 6006-6014.

20   8.   Thornhill, E.M. and Verhoeven, D. (2020) Respiratory Syncytial Virus's Non-structural
21   Proteins: Masters of Interference. Front Cell Infect Microbiol, 10, 225.

22   9.   Bermingham, A. and Collins, P.L. (1999) The M2-2 protein of human respiratory
23   syncytial virus is a regulatory factor involved in the balance between RNA replication
24   and transcription. Proc Natl Acad Sci U S A, 96, 11259-11264.

25   10.   Gould, P.S. and Easton, A.J. (2005) Coupled translation of the respiratory syncytial virus
26   M2 open reading frames requires upstream sequences. J Biol Chem, 280, 21972-21980.

27   11.   Gould, P.S. and Easton, A.J. (2007) Coupled translation of the second open reading
28   frame of M2 mRNA is sequence dependent and differs significantly within the subfamily
29   Pneumovirinae. J Virol, 81, 8488-8496.

30   12.   Phillips, S.L., Soderblom, E.J., Bradrick, S.S. and Garcia-Blanco, M.A. (2016)
31   Identification of Proteins Bound to Dengue Viral RNA In Vivo Reveals New Host
32   Proteins Important for Virus Replication. mBio, 7, e01865-01815.

33   13.   Szutkowska, B., Wieczorek, K., Kierzek, R., Zmora, P., Peterson, J.M., Moss, W.N.,
34   Mathews, D.H. and Kierzek, E. (2022) Secondary Structure of Influenza A Virus
35   Genomic Segment 8 RNA Folded in a Cellular Environment. Int J Mol Sci, 23.

29

1   14.   Michalak, P., Piasecka, J., Szutkowska, B., Kierzek, R., Biala, E., Moss, W.N. and
2   Kierzek, E. (2021) Conserved Structural Motifs of Two Distant IAV Subtypes in
3   Genomic Segment 5 RNA. Viruses, 13.

4   15.   Priore, S.F., Moss, W.N. and Turner, D.H. (2012) Influenza A virus coding regions
5   exhibit host-specific global ordered RNA structure. PLoS One, 7, e35989.

6   16.   Michalak, P., Soszynska-Jozwiak, M., Biala, E., Moss, W.N., Kesy, J., Szutkowska, B.,
7   Lenartowicz, E., Kierzek, R. and Kierzek, E. (2019) Secondary structure of the segment 5
8   genomic RNA of influenza A virus and its application for designing antisense
9   oligonucleotides. Sci Rep, 9, 3801.

10 17.   Simon, L.M., Morandi, E., Luganini, A., Gribaudo, G., Martinez-Sobrido, L., Turner,
11   D.H., Oliviero, S. and Incarnato, D. (2019) In vivo analysis of influenza A mRNA
12   secondary structures identifies critical regulatory motifs. Nucleic Acids Res, 47, 7003-
13   7017.

14 18.   Baranovskaya, I., Sergeeva, M., Fadeev, A., Kadirova, R., Ivanova, A., Ramsay, E. and
15   Vasin, A. (2019) Changes in RNA secondary structure affect NS1 protein expression
16   during early stage influenza virus infection. Virol J, 16, 162.

17 19.   Shafiuddin, M. and Boon, A.C.M. (2019) RNA Sequence Features Are at the Core of
18   Influenza A Virus Genome Packaging. J Mol Biol, 431, 4217-4228.

19 20.   Rodriguez, L., Cuesta, I., Asenjo, A. and Villanueva, N. (2004) Human respiratory
20   syncytial virus matrix protein is an RNA-binding protein: binding properties, location and
21   identity of the RNA contact residues. J Gen Virol, 85, 709-719.

22 21.   Richard, C.A., Rincheval, V., Lassoued, S., Fix, J., Cardone, C., Esneau, C., Nekhai, S.,
23   Galloux, M., Rameix-Welti, M.A., Sizun, C. et al. (2018) RSV hijacks cellular protein
24   phosphatase 1 to regulate M2-1 phosphorylation and viral transcription. PLoS Pathog, 14,
25   e1006920.

26 22.   Gilman, M.S.A., Liu, C., Fung, A., Behera, I., Jordan, P., Rigaux, P., Ysebaert, N.,
27   Tcherniuk, S., Sourimant, J., Eleouet, J.F. et al. (2019) Structure of the Respiratory
28   Syncytial Virus Polymerase Complex. Cell, 179, 193-204 e114.

29 23.   Andrews, R.J., Roche, J. and Moss, W.N. (2018) ScanFold: an approach for genome-
30   wide discovery of local RNA structural elements-applications to Zika virus and HIV.
31   PeerJ, 6, e6136.

32 24.   Gias, E., Nielsen, S.U., Morgan, L.A. and Toms, G.L. (2008) Purification of human
33   respiratory syncytial virus by ultracentrifugation in iodixanol density gradient. J Virol
34   Methods, 147, 328-332.

35 25.   Rameix-Welti, M.A., Le Goffic, R., Herve, P.L., Sourimant, J., Remot, A., Riffault, S.,
36   Yu, Q., Galloux, M., Gault, E. and Eleouet, J.F. (2014) Visualizing the replication of
37   respiratory syncytial virus in cells and in living mice. Nat Commun, 5, 5104.

26. Andrews, R.J., Roche, J. and Moss, W.N. (2018) ScanFold: an approach for genome-wide discovery of local RNA structural elements—applications to Zika virus and HIV. PeerJ, 6, e6136.

27. Katoh, K., Rozewicki, J. and Yamada, K.D. (2017) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics, bbx108-bbx108.

28. Kuraku, S., Zmasek, C.M., Nishimura, O. and Katoh, K. (2013) aLeaves facilitates on-demand exploration of metazoan gene family trees on MAFFT sequence alignment server with enhanced interactivity. Nucleic Acids Res, 41, W22-28.

29. Lusvarghi, S., Sztuba-Solinska, J., Purzycka, K.J., Rausch, J.W. and Le Grice, S.F. (2013) RNA secondary structure prediction using high-throughput SHAPE. J Vis Exp, e50243.

30. Lorenz, R., Bernhart, S.H., Honer Zu Siederdissen, C., Tafer, H., Flamm, C., Stadler, P.F. and Hofacker, I.L. (2011) ViennaRNA Package 2.0. Algorithms Mol Biol, 6, 26.

31. Karabiber, F., McGinnis, J.L., Favorov, O.V. and Weeks, K.M. (2013) QuShape: rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. RNA, 19, 63-73.

32. Johnson, P.Z., Kasprzak, W.K., Shapiro, B.A. and Simon, A.E. (2019) RNA2Drawer: geometrically strict drawing of nucleic acid structures with graphical structure editing and highlighting of complementary subsequences. RNA Biol, 16, 1667-1671.

33. Buchholz, U.J., Finke, S. and Conzelmann, K.K. (1999) Generation of bovine respiratory syncytial virus (BRSV) from cDNA: BRSV NS2 is not essential for virus replication in tissue culture, and the human RSV leader region acts as a functional BRSV genome promoter. J Virol, 73, 251-259.

34. Andrews, R.J., Baber, L. and Moss, W.N. (2017) RNAStructuromeDB: A genome-wide database for RNA structural inference. Sci Rep, 7, 17269.

35. Peterson, J.M., O'Leary, C.A. and Moss, W.N. (2022) In silico analysis of local RNA secondary structure in influenza virus A, B and C finds evidence of widespread ordered stability but little evidence of significant covariation. Sci Rep, 12, 310.

36. Watts, J.M., Dang, K.K., Gorelick, R.J., Leonard, C.W., Bess, J.W., Jr., Swanstrom, R., Burch, C.L. and Weeks, K.M. (2009) Architecture and secondary structure of an entire HIV-1 RNA genome. Nature, 460, 711-716.

37. Low, J.T. and Weeks, K.M. (2010) SHAPE-directed RNA secondary structure prediction. Methods, 52, 150-158.

38. Siegfried, N.A., Busan, S., Rice, G.M., Nelson, J.A. and Weeks, K.M. (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). Nat Methods, 11, 959-965.

39. Deigan, K.E., Li, T.W., Mathews, D.H. and Weeks, K.M. (2009) Accurate SHAPE-directed RNA structure determination. Proc Natl Acad Sci U S A, 106, 97-102.

40. Kutchko, K.M. and Laederach, A. (2017) Transcending the prediction paradigm: novel applications of SHAPE to RNA function and evolution. Wiley Interdisc Rev RNA, 8.

41. Mahen, E.M., Watson, P.Y., Cottrell, J.W. and Fedor, M.J. (2010) mRNA secondary structures fold sequentially but exchange rapidly in vivo. PLoS Biol, 8, e1000307.

42. Bevilacqua, P.C., Ritchey, L.E., Su, Z. and Assmann, S.M. (2016) Genome-Wide Analysis of RNA Secondary Structure. Annu Rev Genet, 50, 235-266.

43. Zeng, M., Duan, Y., Zhang, W., Wang, M., Jia, R., Zhu, D., Liu, M., Zhao, X., Yang, Q., Wu, Y. et al. (2020) Universal RNA Secondary Structure Insight Into Mosquito-Borne Flavivirus (MBFV) cis-Acting RNA Biology. Front Microbiol, 11, 473.

44. Li, Z. and Nagy, P.D. (2011) Diverse roles of host RNA binding proteins in RNA virus replication. RNA Biol, 8, 305-315.

45. Yu, J.M., Fu, Y.H., Peng, X.L., Zheng, Y.P. and He, J.S. (2021) Genetic diversity and molecular evolution of human respiratory syncytial virus A and B. Sci Rep, 11, 12941.

46. Tan, L., Lemey, P., Houspie, L., Viveen, M.C., Jansen, N.J., van Loon, A.M., Wiertz, E., van Bleek, G.M., Martin, D.P. and Coenjaerts, F.E. (2012) Genetic variability among complete human respiratory syncytial virus subgroup A genomes: bridging molecular evolutionary dynamics and epidemiology. PLoS One, 7, e51439.

47. Walsh EE, Perez Marc G, Zareba AM, Falsey AR, Jiang Q, Patton M, Polack FP, Llapur C, Doreski PA, Ilangovan K, Ramet M, Fukushima Y, Hussen N, Bont LJ, Cardona J, DeHaan E, Castillo Villa G, Ingilizova M, Eiras D, Mikati T, Shah RN, Schneider K, Cooper D, Koury K, Lino MM, Anderson AS, Jansen KU, Swanson KA, Gurtman A, Gruber WC, Schmoele-Thoma B, Group RCT. 2023. Efficacy and Safety of a Bivalent RSV Prefusion F Vaccine in Older Adults. N Engl J Med 388:1465-1477.

48. Leroux-Roels I, Davis MG, Steenackers K, Essink B, Vandermeulen C, Fogarty C, Andrews CP, Kerwin E, David MP, Fissette L, Vanden Abeele C, Collete D, de Heusch M, Salaun B, De Schrevel N, Koch J, Verheust C, Dezutter N, Struyf F, Mesaros N, Tica J, Hulstrom V. 2023. Safety and Immunogenicity of a Respiratory Syncytial Virus Prefusion F (RSVPreF3) Candidate Vaccine in Older Adults: Phase 1/2 Randomized Clinical Trial. J Infect Dis 227:761-772.

49. Donovan-Banfield I, Milligan R, Hall S, Gao T, Murphy E, Li J, Shawli GT, Hiscox J, Zhuang X, McKeating JA, Fearns R, Matthews DA. 2022. Direct RNA sequencing of respiratory syncytial virus infected human cells generates a detailed overview of RSV polycistronic mRNA and transcript abundance. PLoS One 17:e0276697.

32

1   50.   Piedra FA, Qiu X, Teng MN, Avadhanula V, Machado AA, Kim DK, Hixson J, Bahl J,
2   Piedra PA. 2020. Non-gradient and genotype-dependent patterns of RSV gene
3   expression. PLoS One 15:e0227558.
4   51.   Wen X, Mousa JJ, Bates JT, Lamb RA, Crowe JE, Jr., Jardetzky TS. 2017. Structural
5   basis for antibody cross-neutralization of respiratory syncytial virus and human
6   metapneumovirus. Nat Microbiol 2:16272.
7   52.   Rincheval V, Lelek M, Gault E, Bouillier C, Sitterlin D, Blouquit-Laye S, Galloux M,
8   Zimmer C, Eleouet JF, Rameix-Welti MA. 2017. Functional organization of cytoplasmic
9   inclusion bodies in cells infected by respiratory syncytial virus. Nat Commun 8:563.
10 53.   Szutkowska B, Wieczorek K, Kierzek R, Zmora P, Peterson JM, Moss WN, Mathews
11   DH, Kierzek E. 2022. Secondary Structure of Influenza A Virus Genomic Segment 8
12   RNA Folded in a Cellular Environment. Int J Mol Sci 23.
13 54.   Michalak P, Soszynska-Jozwiak M, Biala E, Moss WN, Kesy J, Szutkowska B,
14   Lenartowicz E, Kierzek R, Kierzek E. 2019. Secondary structure of the segment 5
15   genomic RNA of influenza A virus and its application for designing antisense
16   oligonucleotides. Sci Rep 9:3801.
17 55.   Lee N, Le Sage V, Nanni AV, Snyder DJ, Cooper VS, Lakdawala SS. 2017. Genome-
18   wide analysis of influenza viral RNA and nucleoprotein association. Nucleic Acids Res
19   45:8968-8977.
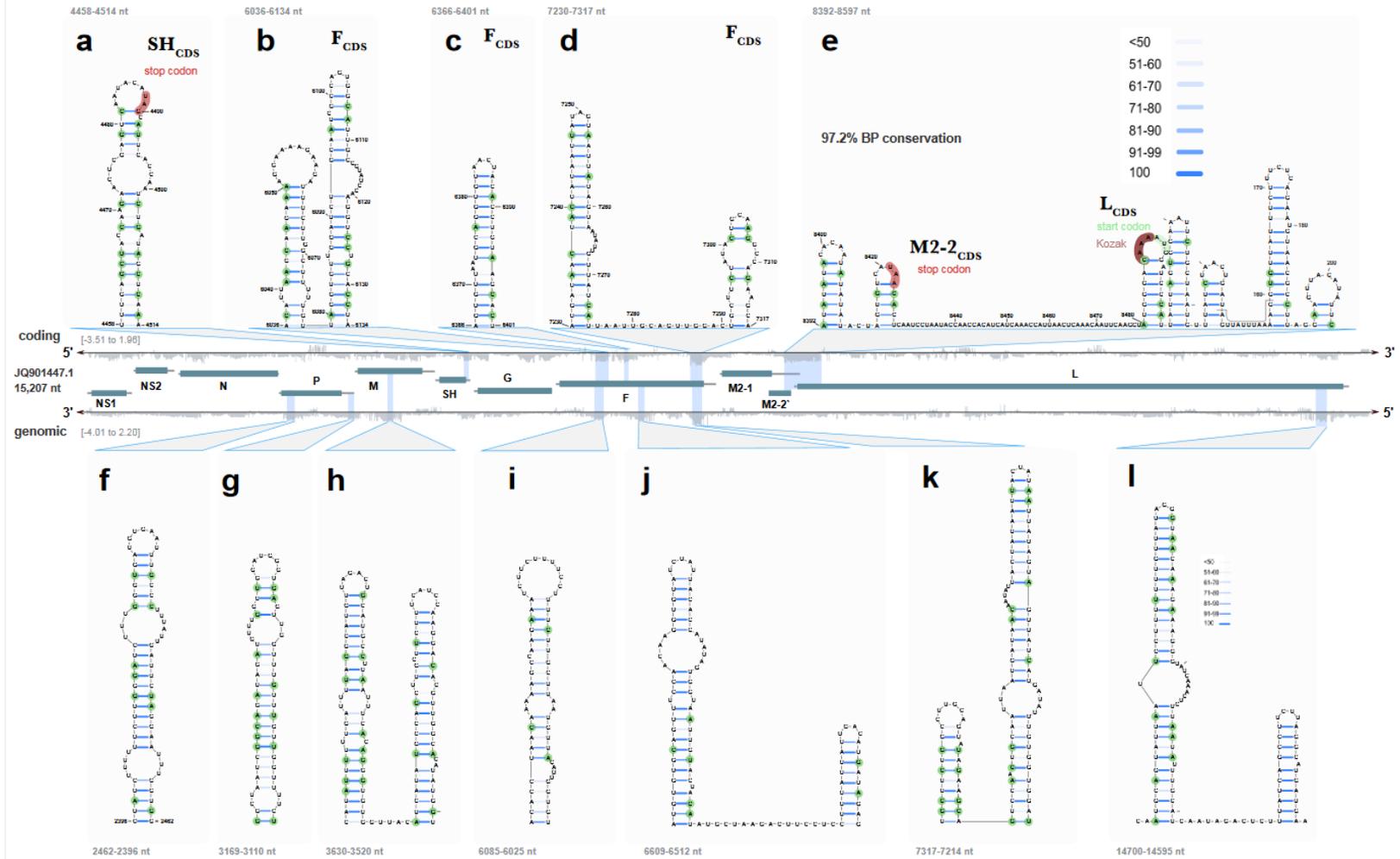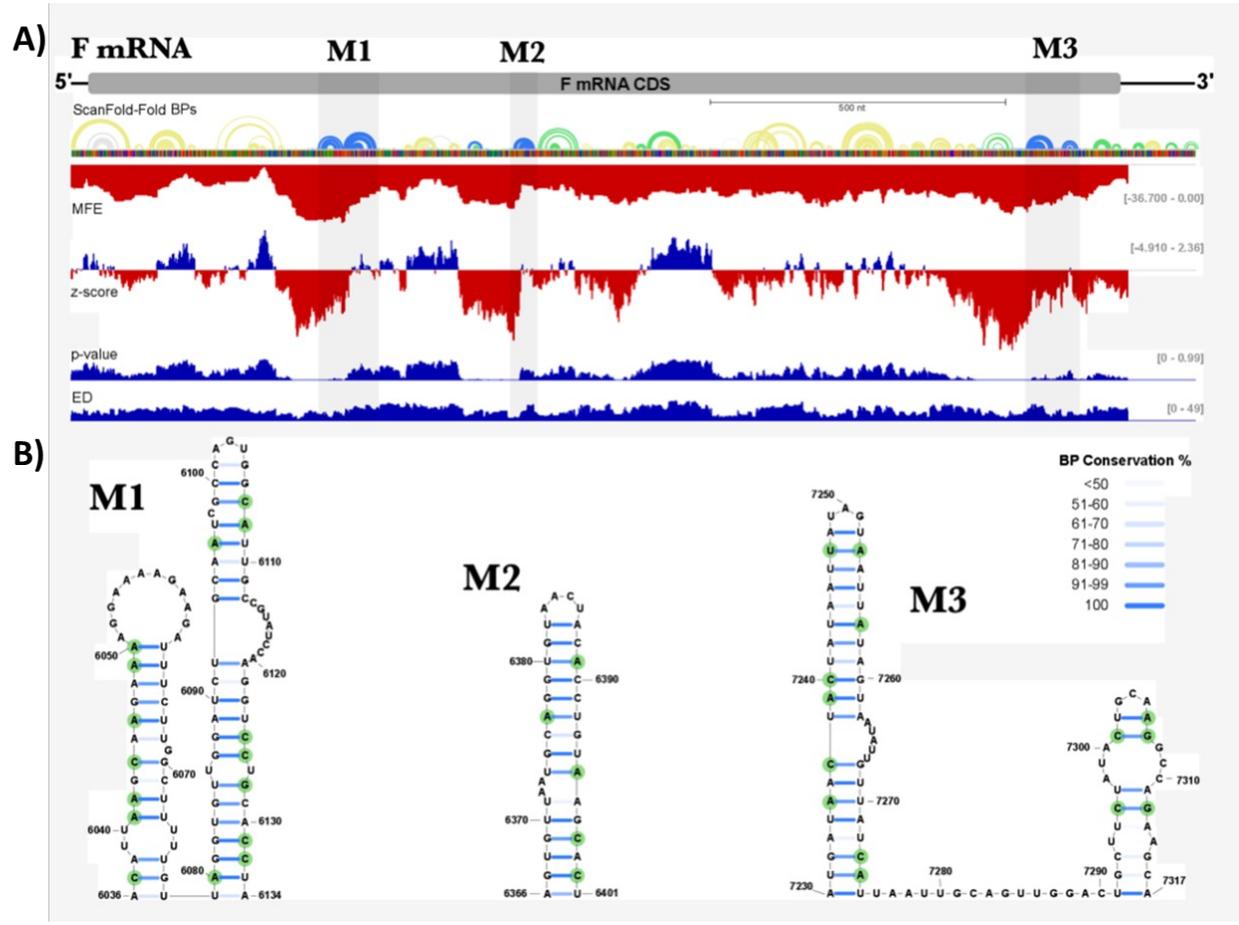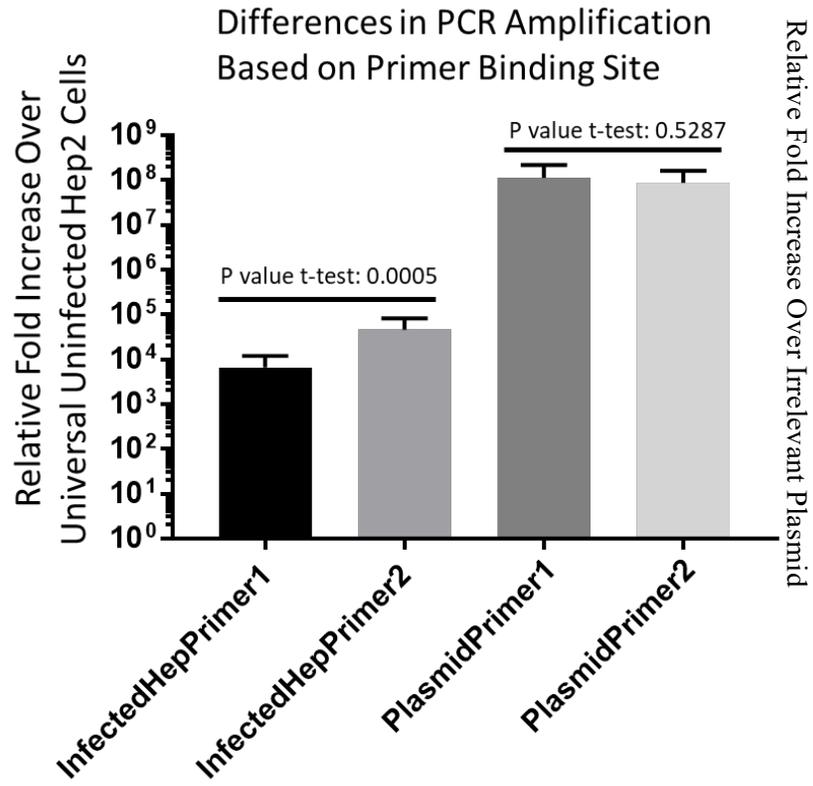
20

Figure 1

Figure 2

A) F mRNA · M1 · M2 · M3 · F mRNA CDS · ScanFold-Fold BPs · MFE · z-score · p-value · ED

B) M1 · M2 · M3 · BP Conservation %

Figure 3

Differences in PCR Amplification
Based on Primer Binding Site
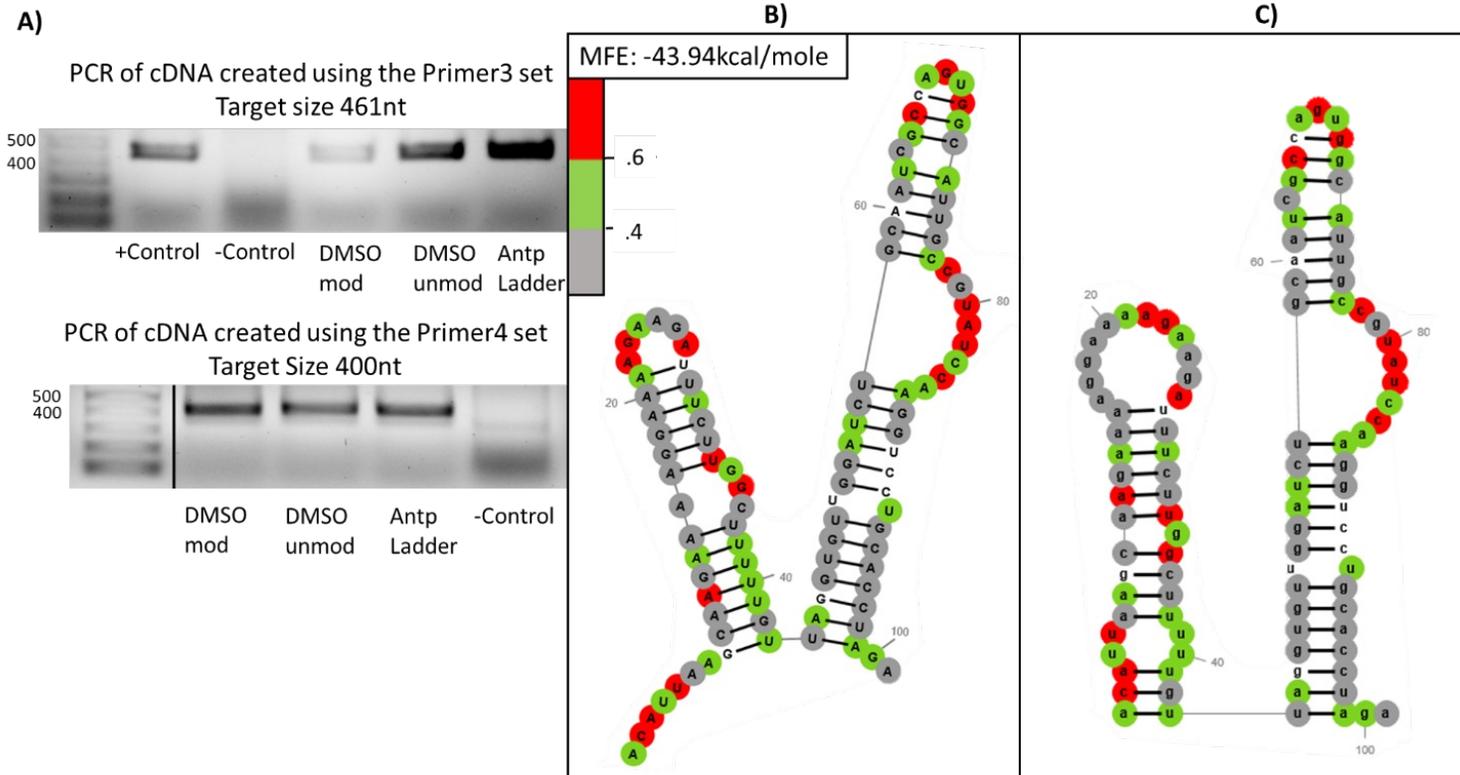
Figure 4

Figure 5

A)
```
WT)  AC ATTAAG C AA GAAAA G GAAAA GAA GATTTCTTGG TTTTTT G
Relx) AC CCTGAG T AA AAAGC G CAAGC G CCGGTTCCTCGG CTTCTT C
```


B)

MFE: -43.94kcal/mole

C)

WT)   Relx)
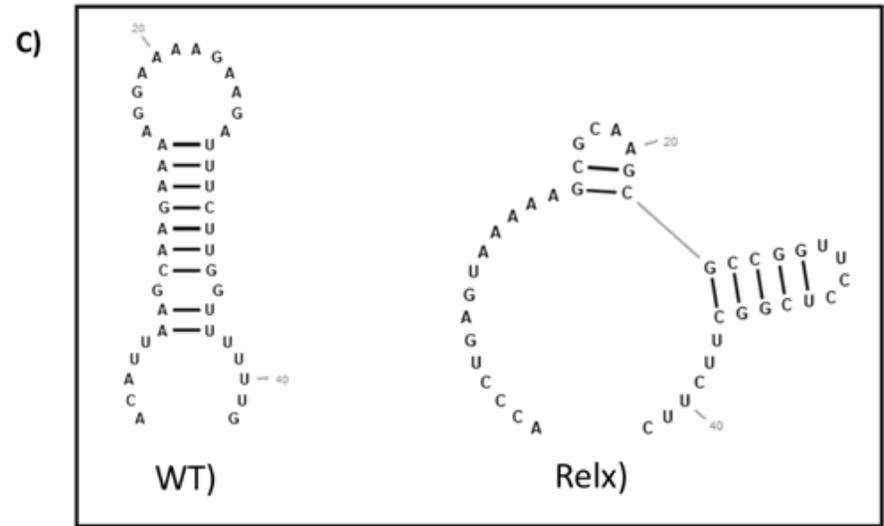
D) Amino Acid Sequence) TLSKKRKRRFLGFL
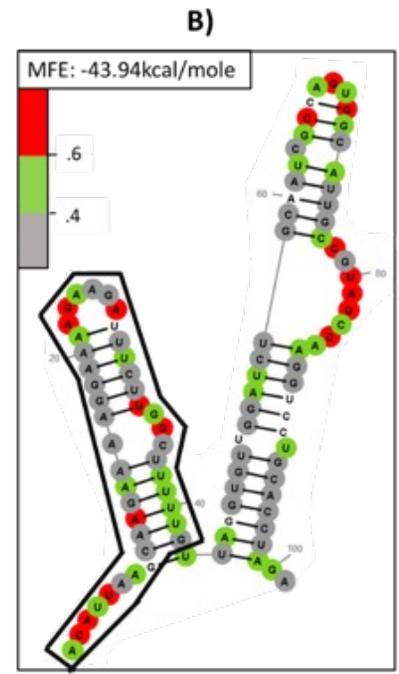
Figure 6
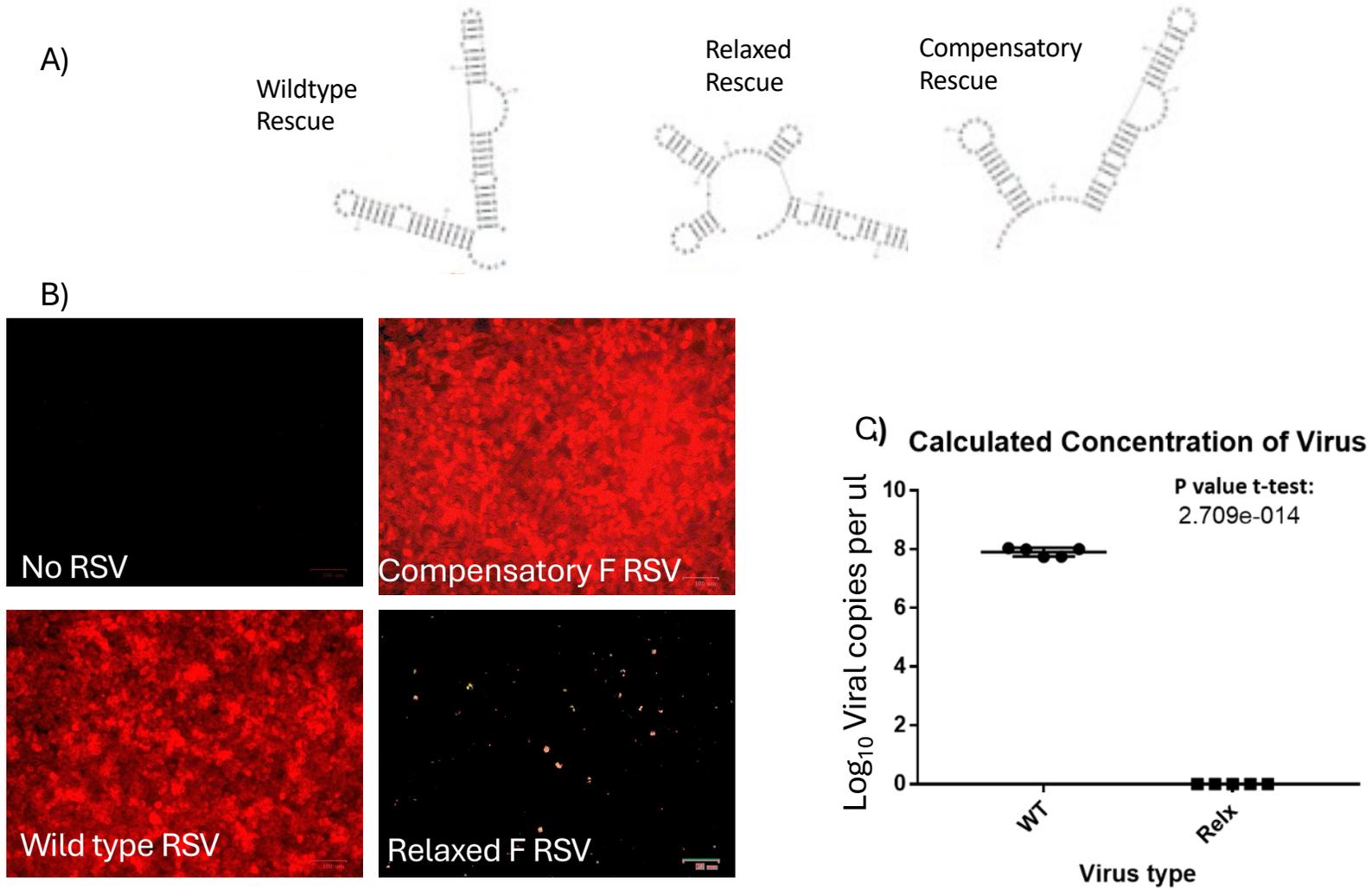
A) **mRNA Concentrations off of pcr tranfections**

B)

Figure 7

A)

Wildtype
Rescue

Relaxed
Rescue

Compensatory
Rescue

No RSV

Compensatory F RSV

Wild type RSV

Relaxed F RSV

Calculated Concentration of Virus

P value t-test:
2.709e-014

Log_{10} Viral copies per

8 —

6 —

4 —

2 —

0 —

WT

Relx

Virus type
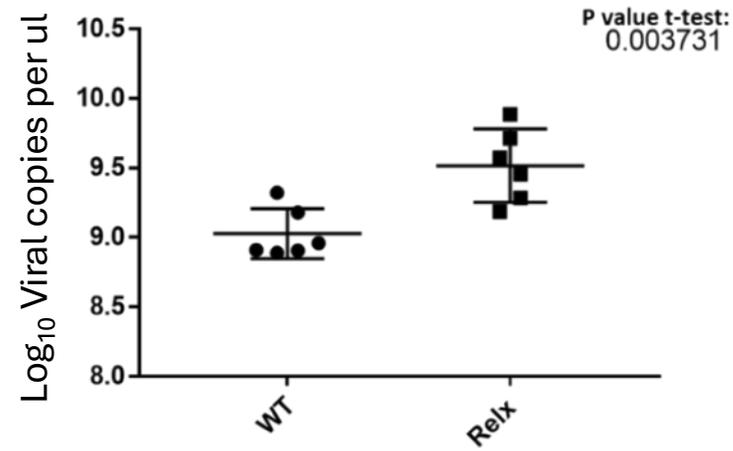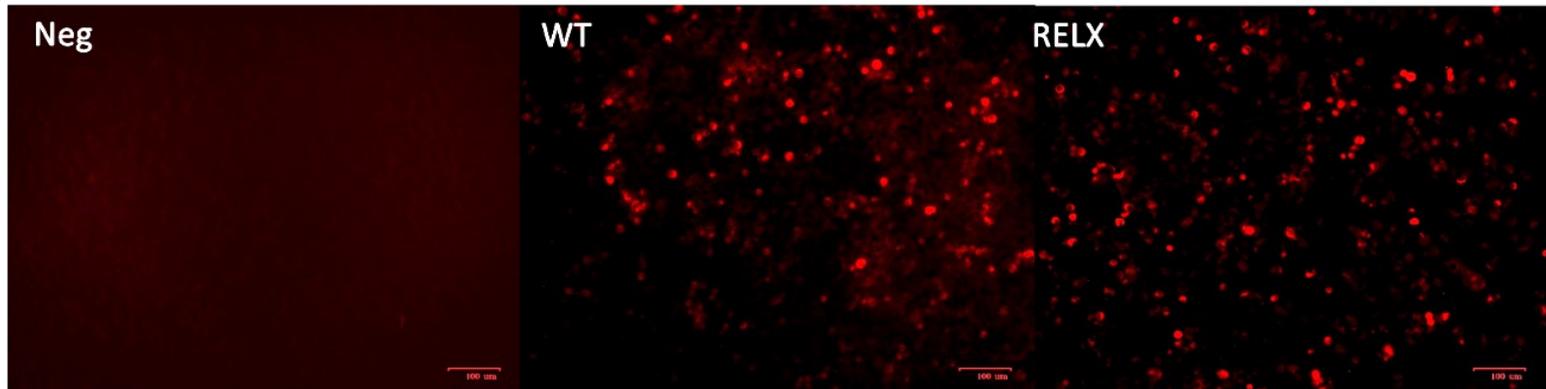
Figure 8

A)

Sequence scan results

Your sequence:

```
1   CCAAAACCAC CAAUGUAACA UUAAGCAAGA AAAGGAAAAG AAGAUUUCUU GGCUUUUUGU UAGGUGUUGG AUCUGCAAUC GCCAGUGGCA UUGCCGUAUC
101 CAAGGUCCUG CACCUAGAAG GGGAAGUGAA CAAAAUCAAA AGUGCUCUAC
```

Score	Relative Score	RBP Name	Start	End	Matching sequence	Matrix ID	Download PWM	Download PFM
7.3693752	100%	FUS	63	66	GGUG	637_11098054	Download PWM	Download PFM
7.2294196	100%	Pum2	14	17	UGUA	329_11780640	Download PWM	Download PFM
6.6279899	100%	MBNL1	143	146	UGCU	669_20071745	Download PWM	Download PFM
6.4668404	100%	EIF4B	122	125	GGAA	352_8846295	Download PWM	Download PFM
6.4668404	100%	EIF4B	34	37	GGAA	352_8846295	Download PWM	Download PFM
6.33985	100%	KHSRP	105	108	GUCC	1186_17893325	Download PWM	Download PFM
5.2682554	100%	RBMX	82	85	CCAG	922_19282290	Download PWM	Download PFM
4.8027008857	93%	SFRS13A	37	43	AAAGAAG	1169_19561594	Download PWM	Download PFM
4.6667232	88%	RBMX	7	10	CCAC	922_19282290	Download PWM	Download PFM
4.62028767	100%	SFRS1	33	36	AGGA	1173_19561594	Download PWM	Download PFM
4.50571335	88%	SFRS13A	31	37	AAAGGAA	1169_19561594	Download PWM	Download PFM
4.1881759	90%	SFRS1	68	71	UGGA	1173_19561594	Download PWM	Download PFM
3.82636396	86%	ELAVL1	44	47	AUUU	1170_19561594	Download PWM	Download PFM

B)