

Article

Secondary Structure of Subgenomic RNA M of SARS-CoV-2

Marta Soszynska-Jozwiak¹, Agnieszka Ruskowska¹ , Ryszard Kierzek¹, Collin A. O'Leary² , Walter N. Moss² and Elzbieta Kierzek^{1,*} 

- ¹ Institute of Bioorganic Chemistry, Polish Academy of Sciences, 61-704 Poznan, Poland; mjozwiak@ibch.poznan.pl (M.S.-J.); aruskowska@ibch.poznan.pl (A.R.); rkierzek@ibch.poznan.pl (R.K.)
² Roy J. Carver Department of Biophysics, Biochemistry and Molecular Biology, Iowa State University, Ames, IA 50011, USA; caoleary@iastate.edu (C.A.O.); wmos@iastate.edu (W.N.M.)
* Correspondence: elzbieta.kierzek@ibch.poznan.pl

Abstract: SARS-CoV-2 belongs to the *Coronavirinae* family. Like other coronaviruses, SARS-CoV-2 is enveloped and possesses a positive-sense, single-stranded RNA genome of ~30 kb. Genomic RNA is used as the template for replication and transcription. During these processes, positive-sense genomic RNA (gRNA) and subgenomic RNAs (sgRNAs) are created. Several studies presented the importance of the genomic RNA secondary structure in SARS-CoV-2 replication. However, the structure of sgRNAs has remained largely unsolved so far. In this study, we probed the sgRNA M model of SARS-CoV-2 in vitro. The presented model molecule includes 5'UTR and a coding sequence of gene M. This is the first experimentally informed secondary structure model of sgRNA M, which presents features likely to be important in sgRNA M function. The knowledge of sgRNA M structure provides insights to better understand virus biology and could be used for designing new therapeutics.

Keywords: RNA structure; SARS-CoV-2; COVID-19; subgenomic RNAs; sgRNA; sgRNA M; chemical mapping



Citation: Soszynska-Jozwiak, M.; Ruskowska, A.; Kierzek, R.; O'Leary, C.A.; Moss, W.N.; Kierzek, E. Secondary Structure of Subgenomic RNA M of SARS-CoV-2. *Viruses* **2022**, *14*, 322. <https://doi.org/10.3390/v14020322>

Academic Editors: Célia F. Rodrigues and Natália Cruz-Martins

Received: 10 December 2021

Accepted: 31 January 2022

Published: 4 February 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) causes coronavirus disease 19 (COVID-19) and is responsible for widespread infection/death and concomitant disruptions to health services, travel, trade, education and has a negative impact on people's physical and mental health [1]. SARS-CoV-2 belongs to the Betacoronavirus genus and is a member of the *Coronaviridae* family, which also includes alpha-, gamma- and deltacoronaviruses [2]. Like other coronaviruses, SARS-CoV-2 is enveloped and possess three structural proteins: membrane protein (M), spike protein (S), and envelope protein (E), while nucleocapsid protein (N) protects the viral RNA genome by forming a capsid [3]. SARS-CoV-2 also produces sixteen non-structural proteins (nsp1–16) and accessory proteins [4]. SARS-CoV-2 is an RNA virus and possesses a positive-sense, single-stranded RNA genome of ~30 kb [5]. The genomic RNA has a 5' cap and 3' polyA tail and is used for the production of two large overlapping polyproteins (pp): pp1a and pp1ab. Polyproteins pp1a and pp1ab contain the non-structural proteins 1–11 and 1–16, respectively. Many of them, with an N protein, create the replicase–transcriptase complex (RTC) [6]. Genomic RNA is used as the template for replication and transcription by RTC. These processes result in the generation of negative-sense RNA intermediates that serve as templates for the production of positive-sense genomic RNA (gRNA) and subgenomic RNAs (sgRNAs). The gRNA is packaged into progeny virions or is used for translation, while sgRNAs encode conserved structural proteins, nucleocapsid proteins and several accessory proteins [6–9].

In coronaviruses, each sgRNA possesses a short 5'-terminal leader sequence derived from the 5' end of the genome. Transcription regulatory sequences (TRS) are necessary to add the leader sequence to sgRNAs [10]. TRSs are located at the 3' end of the leader

sequence (leader TRS, TRS-L), as well as upstream of the genes in the 3'-proximal part of the genome (body TRSs, TRS-B). TRSs contain a conserved 6–7 nt core sequence (CS) surrounded by variable sequences. During negative-strand synthesis, *RNA-dependent RNA polymerase* (RdRP) pauses when it crosses a body TRS and switches the template to the leader TRS. As a result, sgRNA possess a leader sequence derived from the 5' untranslated region of the genome and a TRS 5' of the open reading frame [11,12].

RNA secondary structure in untranslated and coding regions play a key role in the viral replication cycle [13]. The secondary structure within the RNA of SARS-CoV-2 has been intensively studied. Bioinformatic analysis showed that the SARS-CoV-2 genome has almost twice the propensity to form a secondary structure than one of the most structured RNA genomes in nature, the HCV genome [14,15]. Recently, detailed secondary structure models for the extended 5'UTR, frameshifting stimulation element, 3'UTR and regions of the SARS-CoV-2 viral genome that have a high propensity for RNA secondary structures and are conserved within SARS-CoV-2 strains were predicted [15,16] SARS-CoV-2 was also investigated for the presence of large-scale internal RNA base pairing (genome-scale ordered RNA structure) (GORS) in its genome. This analysis showed the existence of 657 stem-loop structures and 2015 duplexes [17]. Data revealed that regions containing the highest amount of structure within the SARS-CoV-2 genome are in the 5' end, as well as regions corresponding to glycoproteins S and M [18]. Recently, RNA structure probing of the full SARS-CoV-2 coronavirus genome both in vitro and in living infected cells was published [19–23]. Additionally, using RNA structure probing with nanopore direct-RNA sequencing, NAI reactivity values for 3a, E, M, 6, 7a, 7b, 8 and N sgRNAs were measured [23]. However, the structure model of sgRNA M has not yet been determined and discussed.

In this study, we performed probing of the sgRNA M sequence of SARS-CoV-2 in vitro. We used selective 2'-hydroxyl acylation analyzed by a primer extension (SHAPE) method and chemical mapping with dimethyl sulfate (DMS) and 1-cyclohexyl-(2-morpholinoethyl) carbodiimide metho-p-toluene sulfonate (CMCT) to obtain the secondary structure of sgRNA M. This is the first experimentally informed secondary structure model of sgRNA M, which presents features likely to be important in sgRNA M function. Several chemical reagents combined with a bioinformatics program allowed for the prediction of a high confidence structure. The comprehensive characterization of sgRNA M structure will provide important insights to better understand virus biology and could be used for designing new therapeutics.

2. Material and Method

2.1. Experimental Constructs

The DNA template for the synthesis of sgRNA M was obtained in several steps. Firstly, reverse transcription was carried out using SuperScript III (Thermo Fisher Scientific) with Random Primers Mix (New England BioLabs, Ipswich, MA, USA) on RNA of SARS-CoV-2 from strain Slovakia/SK-BMC5/2020 (received from <https://www.european-virus-archive.com>, accessed on: 9 July 2020). Next, three PCR reactions using cDNA as template and specific primers (F1 and RM, F2 and RM, F3 and RM, Table 1) were performed to amplify the TRS-M coding sequence and add the leader sequence and transcription promoter on the 5' end of sgRNA M. After this step, primers FC and RC were used to add an EcoRI site on the 5' end and a Pst I site on the 3' end of the template of sgRNA M. DNA was purified using the Pure Link™ PCR Micro Kit (Thermo Fisher Scientific, Waltham, MA, USA). The DNA template was cloned into pUC19 and sequenced using the M13F and M13R primers for confirmation of proper sequence (Table 1).

Table 1. Primers for polymerase chain reaction used to obtain DNA template for sgRNA M. The underlined nucleotide residues are the polymerase T7 promoter.

Primer Name	Primer Length (nt)	Sequence 5'→3'
FM	46	GCGT <u>AATACGACTCACTATAGGG</u> ATTAAGGTTTATACCTTCCCAG
RM	29	TTACTGTACAAGCAAAGCAATATTGTCAC
F1	54	CTTGTAGATCTGTTCTCTAAACGAACTAAATATTATATTAGTTTTTCTGTTGG
F2	48	GTAACAAACCAACCAACTTTTCGATCT CTTGTAGATCTGTTCTCTAAAC
F3	45	ATTAAGGTTTATACCTTCCCAGGTAACAAACCAACCAACTTTCG
FC	29	TTCTGCAG ATTAAGGTTTATACCTTCCC
RC	37	ATGAATTCTTACTGTACAAGCAAAGCAATATTGTCAC
M13F	24	CGCCAGGGTTTTCCCAGTCACGAC

2.2. Oligonucleotides Synthesis and Labelling

Primers for reverse transcription were synthesized by the phosphoramidite approach on a MerMade synthesizer. Primers for reverse transcription were synthesized with fluorophores: 6-carboxyfluorescein (6-FAM) and 5-carboxy-4',5'-dichloro-2',7'-dimethoxyfluorescein (5-JOE) on the 5'-end (Table 2). Primers were deprotected and purified according to published protocols [24,25]. Concentrations of all oligonucleotides were measured using a Spectrophotometer UV (NanoDrop2000 Thermo Fisher Scientific, Waltham, MA, USA). Primers for PCR were purchased from Sigma-Aldrich, Saint Louis, MO, USA).

Table 2. Primers for reverse transcription. Each primer was labeled with 6-FAM or 5-JOE at the 5' end.

Primer Name	Primer Length (nt)	Sequence 5'→3'
M1	27	TTACTGTACAAGCAAAGCAATATTGTC
M2	21	CAGCTCCGATTACGAGTTCAC
M3	24	CAAGCTAAAGTTACTGGCCATAAC

2.3. RNA Synthesis

The DNA template for in vitro transcriptions of sgRNA M was obtained by PCR from a modified puC19 plasmid using primers FM and RM (Table 1). DNA was purified using the Pure Link™ PCR Micro Kit (Thermo Fisher Scientific, Waltham, MA, USA). The in vitro transcription reaction was performed using a MEGAscript™ T7 Transcription Kit (Thermo Fisher Scientific) according to the manufacturer's protocol. RNA product was purified using RNeasy MiniElute Cleanup Kit (Qiagen, Hilden, Germany). The integrity and purity of samples were checked on an agarose gel.

2.4. RNA Folding

Before each experiment, RNA was folded in the same manner. RNA was heated to 80 °C in water for 5 min and slowly cooled to 50 °C. At this temperature, folding buffer was added, and samples were slowly cooled to 37 °C. The final concentration of buffer was 300 mM NaCl, 5 mM MgCl₂, 50 mM HEPES, pH 7.5. RNA integrity and homogeneity after folding were analyzed by native gel electrophoresis using 0.8% agarose gel running at 4 °C with low voltage. Under these conditions, one band was observed (Figure S1, Supporting_Information_1).

2.5. Chemical Mapping Using NMIA, DMS and CMCT

The folding of RNA was carried out as described above. Next, chemical mapping was conducted according to published procedures with appropriate optimizations [26–28]. Briefly, 5.6 mM of NMIA, 30 mM of CMCT or 0.18% of DMS were used in mapping reactions. Chemical mapping was performed at 37 °C with DMS, CMCT or NMIA for 15, 30 or 40 min, respectively. Parallel, control reactions were facilitated in the same condition but without mapping reagents. Modified nucleotides were read-out by primer extension using a stoichiometry of 2 pmol primer/2 pmol RNA. Primer extension was performed at 55 °C with reverse transcriptase SuperScript III (Thermo Fisher Scientific) using the manufacturer's protocol. Next, cDNA fragments and ddNTP ladders were separated by capillary electrophoresis (Laboratory of Molecular Biology Techniques at Adam Mickiewicz University in Poznan). Primers labeled with 6-FAM were used for the detection of modification by DMS, CMCT or NMIA and with control reactions without mapping reagents. Samples were resolved in two capillaries (reaction and control) with ddNTP ladders. Primers labeled with 5-JOE were used for ddNTP ladders (most often ddATP). The experiments were performed in at least technical triplicate with the average results presented. To obtain the reactivity values of each nucleotide, the standard deviation (SD) was calculated (Table S1, Supporting_Information_2).

2.6. Processing of Chemical Mapping Data

The QuShape program was used to analyze mapping data according to a published method [29]. NMIA reactivities were normalized by the QuShape program using model-free statistics to a scale spanning 0 to ~2, where zero indicates no reactivity and 1.0 is the low average intensity for highly reactive RNA positions [29]. Nucleotides had normalized SHAPE reactivities 0–0.5, 0.5–0.7, and ≥ 0.7 correspond to unreactive, moderately reactive and highly reactive positions, respectively. Nucleotides with no data were designated as –999. Normalized SHAPE reactivities from the extension reaction of each primer were processed independently. DMS and CMCT modifications analysis was conducted similar to NMIA reactivity calculations, except that only strong modifications (reactivities ≥ 0.7) were used in RNAstructure program prediction.

Chemical mapping results were used in the RNAstructure program [30] for the prediction of the secondary structure of sgRNA M. Normalized SHAPE reactivity (as described above) were used in RNAstructure 6.2 through “Read SHAPE reactivity—pseudo free energy” mode with a slope of 1.8 and intercept of -0.6 kcal/mol [31]. DMS and CMCT strong reactivities were introduced in the same prediction using the “chemical modification” mode [32].

2.7. Bioinformatic Analysis of Base Pairs Probabilities

The sgRNA M base pair probabilities were obtained using the “Partition Function RNA” mode implemented in the RNAstructure program. SHAPE and chemical mapping experiment results were incorporated as constraints after loading the sequence file in “Partition Function RNA” mode, and a .pfs file was generated. All constraints were obtained as described in the *Processing of chemical mapping data* section and were the same as applied for sgRNA M folding. Next, the secondary structure of the sgRNA M model was annotated using the .pfs file using the “Add Probability Color Annotation” mode in the RNAstructure program, version 6.2

2.8. Covariation Analysis

The sequence of the in vitro probed sgRNA M underwent covariation analysis via the cm-builder pipeline, and details of this process are available [19]. Briefly, cm-builder utilizes the programs INFERENCE of RNA ALIGNMENT (INFERNAL) (here, release 1.1.2) [33,34] and R-scape (here, version 1.5.16) [35,36] to make alignments of a reference sequence to homologous sequences and then cross-evaluate a structural model for statistically significant covariation that maintains base pairing. The sgRNA M sequences were aligned to a

previously generated fasta file [15] of 25,571 *Coronaviridae* sequences, obtained from the Virus Pathogen Database and Analysis Resource (ViPR database, <https://www.viprbrc.org/brc/home.spg?decorator=vipr>, accessed on 10 February 2021) [37,38].

Additionally, the sgRNA M sequence was queried against the nucleotide BLAST (Basic Local Alignment Search Tool, <https://blast.ncbi.nlm.nih.gov/Blast.cgi>, accessed on 10 December 2021) database and yielded 453 homologous sequences. These sequences were subsequently MAFFT (Multiple Alignment using Fast Fourier Transform) [39] aligned with the sgRNA M sequence. From here, alignments were used to calculate the conservation of nucleotides in each base pair of the model structure.

3. Results and Discussion

3.1. Structure Probing of sgRNA M of SARS-CoV-2

The sequence of sgRNA M of SARS-CoV-2 was obtained by adding the leader sequence to the M coding sequence from the SARS-CoV-2 strain Slovakia/SK-BMC5/2020 using PCR reactions. The sgRNA M sequence of our model is identical to the SARS-CoV-2 Wuhan-Hu-1 isolate (ID: NC_045512.2). A leader sequence is characteristic for the sgRNA of the Coronaviridae family, and “leader to body fusion” takes place during discontinuous transcription [12]. Chemical mapping was used to determine a secondary structure of sgRNA M. In vitro transcribed sgRNA M was folded in folding buffer (300 mM NaCl, 5 mM MgCl₂, 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.5) to obtain a single RNA conformation, as assessed by non-denaturing agarose gel (Figure S1, Supporting Information 1). Chemical mapping was performed at 37 °C with DMS (methylates N1 of A and N3 of C when unpaired), CMCT (modifies N3 of U and N1 of G when unpaired) and SHAPE reagent N-methylisatoic anhydride (NMIA) (modifies flexible 2'-hydroxyl groups on the ribose) [26–28]. The modifications from chemical mapping were analyzed by reverse transcription followed by capillary electrophoresis (Figures S2 and S3, Supporting Information 1; Table S1, Supporting Information 2).

DMS strongly modified 123 nucleotides (nt), meaning that 32.6% of all adenosines and cytidines in sgRNA M were structurally accessible. CMCT modified 106 nts, which represents 40.9% of all uridines. The SHAPE reagent NMIA strongly modified 119 nts, and 56 nts had moderate modifications signals, representing 22.2% of all sgRNA M nucleotides (see Section 2.6). Chemical mapping data and SHAPE data are complementary to each other and together highlight the eight most flexible regions in sgRNA M (each containing at least seven flexible nucleotides in the sequence): 36–43, 76–86, 260–267, 469–481, 487–494, 596–603, 612–621, 729–743 (Table S1, Supporting Information 2).

3.2. Base Pair Probabilities

To assess prediction quality and identify well-defined structural regions, we calculated the secondary structure partition function using RNAstructure 6.2 and, from this, determined the base pair probabilities for model pairs [40]. For the partition function calculations, experimental data were included (see Section 2.7 for details). Results indicate that there are several regions with paired and unpaired nucleotides of more than 90% probability: 1–121, 131–210, 220–502, 707–766. Additionally, all single-stranded regions are well defined by having a low probability of pairing (Figure 1).

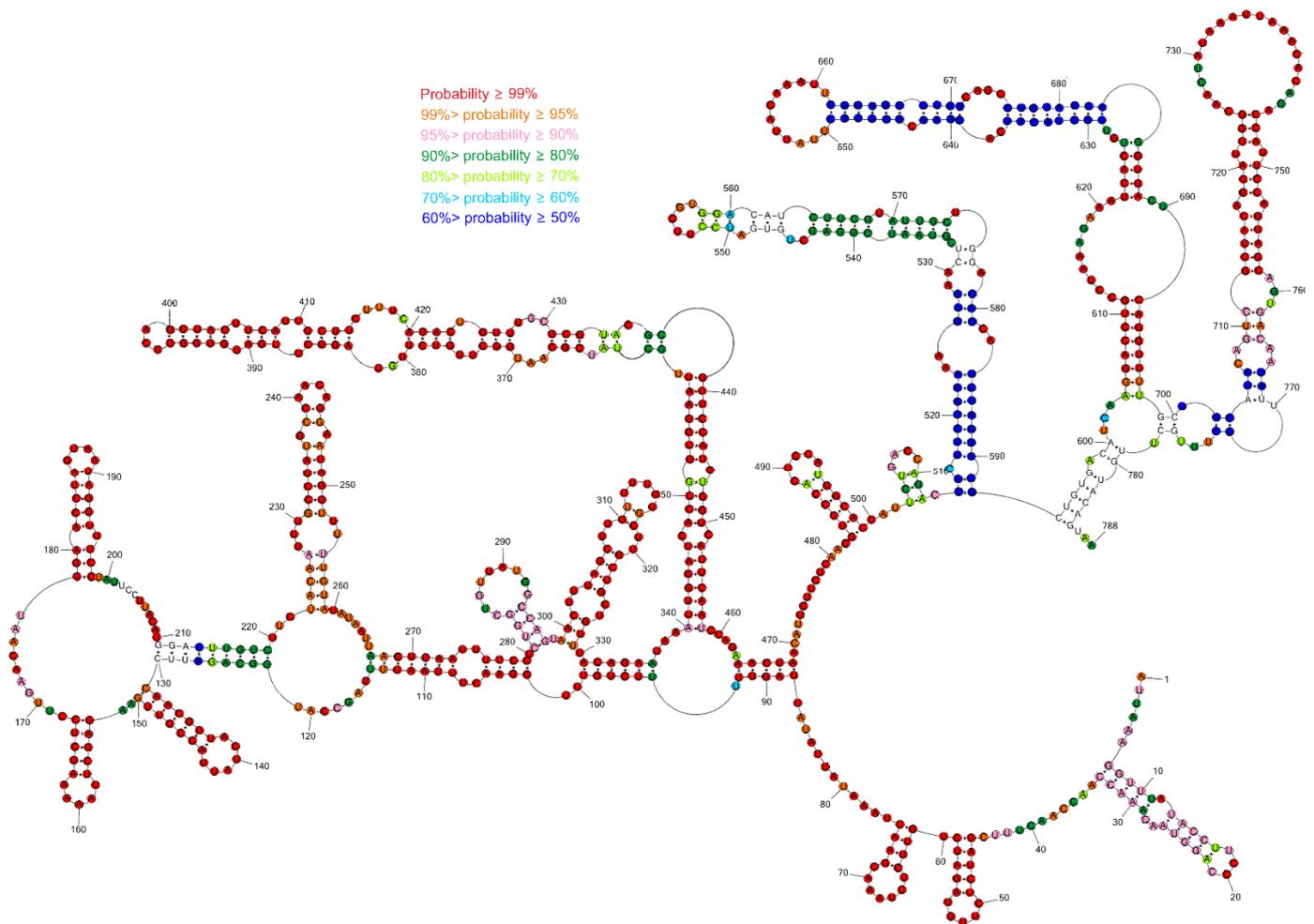


Figure 1. Predicted probability of nucleotides being paired or single-stranded in sgRNA M using the RNAstructure program. Probability lower than 50% is not colored. The partition function calculation incorporated restraints from strong reactivity of DMS and CMCT as well as SHAPE reactivities converted to pseudo-energies.

3.3. Model of Secondary Structure for sgRNA M

To predict the secondary structure of sgRNA M based on the experimental probing data, the results of chemical mapping were used to constrain predictions in the RNAstructure 6.2 program. SHAPE data were loaded as pseudoenergy constraints (the energy contribution of SHAPE reactive nt were penalized) and DMS and CMCT modifications were included as chemical mapping constraints (highly reactive nts are forbidden to be in Watson–Crick base pairs flanked by Watson–Crick base pairs). The default values for slope and intercept in the RNAstructure 6.2 program were used. The default values of these parameters were determined by optimizing the accurate modelling of the SHAPE data set with sequences of known structures [31].

Our model of sgRNA M is highly structured with plenty of accessible bulges and loops (Figure 2). RNA motifs in the sgRNA M model are thermodynamically stable and have high calculated base pair probability. The ΔG°_{37} of the entire folded secondary structure is -416 kcal/mol. We showed that most of the inaccessible regions defined by chemical mapping correspond to areas containing base pairs. The 5' end of the sgRNA M model was folded into three hairpins: SL1, SL2 and SL3. These three hairpins also occur in the 5'UTR of SARS-CoV-2 in its 5' 300 nt fragment [20,41,42] and are present in in vitro models and in-cell models of the whole genome [20–23,43]. These hairpins are also in good agreement with a structural–phylogenetic analysis of group IIb coronaviruses [44] and

in silico prediction of the whole SARS-CoV-2 genome [16,17]. Moreover, the folding of a 5' leader sequence of sgRNA M is in agreement with a study of the secondary structure of sgRNA N [14]. This investigation showed that the 5' leader sequence folds almost autonomously in the sgRNA N, with the exception of a few poorly determined long-range interactions [14]. SL1 is the most variable among SARS-CoV-2 variants [41], generally possessing mismatches, bulges and a high number of A–U and U–A base pairs. This fact causes less thermodynamic stability of SL1 than SL2 and SL3. On the other hand, this feature is important for the replication of mouse hepatitis virus (MHV), a well-studied member of the *Coronaviridae* family [45]. SL2 is conserved in all CoVs, typically containing a pentaloop stacked on a five base-pair stem and creating a U-turn motif. This hairpin plays a critical role in MHV replication and translation [46]. SL3 is conserved only in subgroups of beta and gammaCoVs [4] and contains TRS-L sequences that take part in discontinuous transcription [11,44].

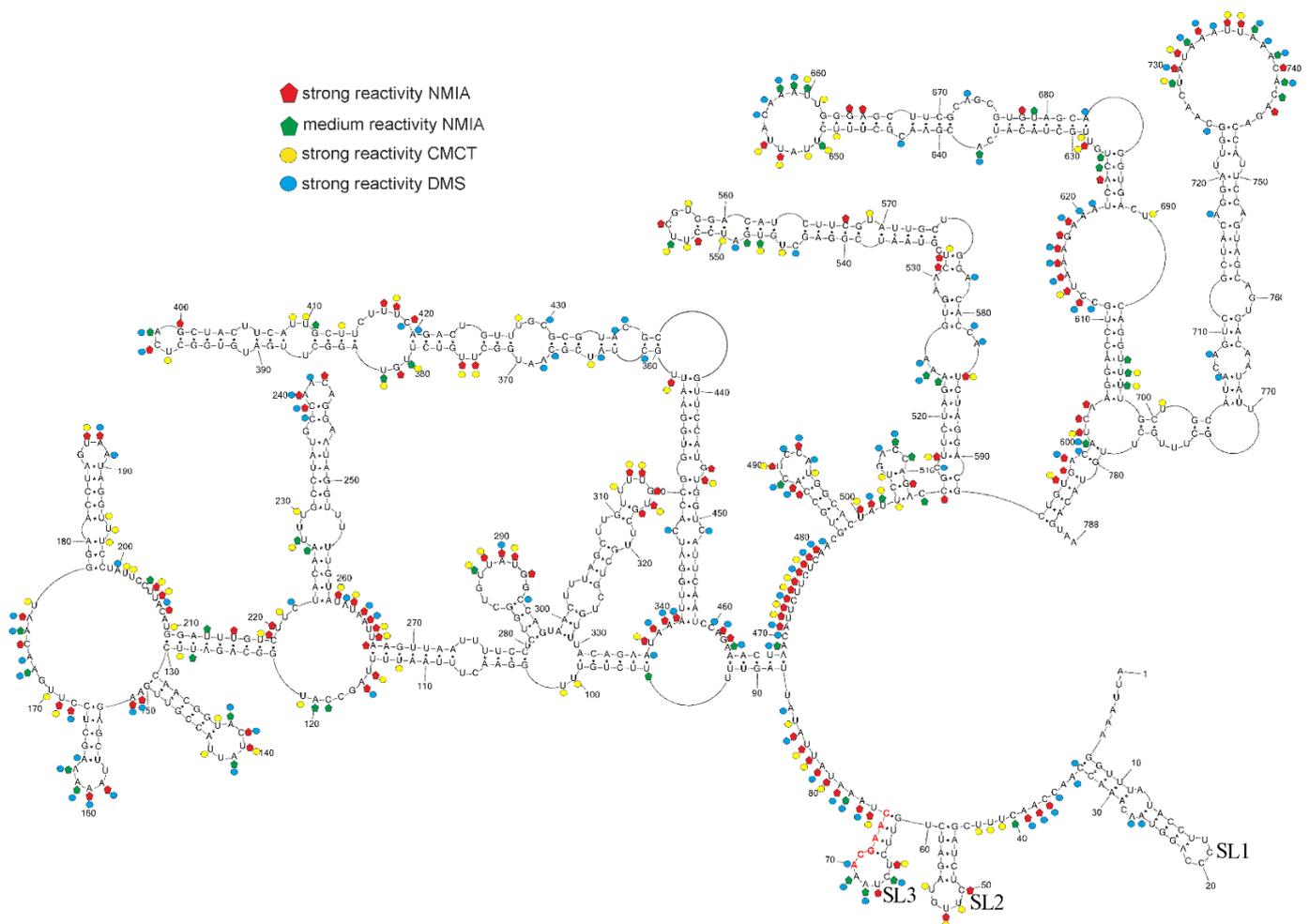


Figure 2. sgRNA M model predicted by RNAstructure 6.2 using experimental data as constraints. Strong DMS and CMCT modifications, as well as SHAPE reactivities converted to pseudo-free energies, were used. The numbering of sgRNA M is from its 5' end. The AUG start codon spans nucleotides 120–122. Red nucleotides indicate TRS sequences. Hairpins SL1, SL2 and SL3, are indicated.

Recently, a prediction of interaction between the SARS-CoV-2 genome and the human proteome indicated that a highly structured region at the 5' end had a large number of interactions with proteins such as (1) ATP-dependent RNA helicase—DDX1, which was previously reported to be essential for Avian infectious bronchitis coronavirus replication [18,47], (2) adenosine deaminases acting on RNA (ADAR) that catalyzes the hydrolytic deamination of adenosine to inosine, which affects viral protein synthesis, proliferation

and infectivity [18,48], and (3) 2'-5'-oligoadenylate synthetases which control viral RNA degradation [18,49,50]. Some of these proteins could interact with a leader sequence of sgRNA. This assumption was confirmed via experiments with DDX1 knockdown that reduced the number of sgRNA in SARS-CoV-1 infected cells [51]. This finding and the preservation of 5'UTR motifs in sgRNA M indicate similar interactions could occur with sgRNA M. Moreover, interactions between the SARS-CoV-2 genome, as well as sgRNAs and host RNAs, were revealed. However, the SARS-CoV-2 genome and sgRNAs take part in different interactions with host RNAs [23,43].

RNA structure probing coupled with nanopore direct-RNA sequencing were used to map sgRNAs with NAI in living cells, but the structure of sgRNA M was not proposed [23]. NMIA reactivities of sgRNA M mapped in vitro obtained herein were compared with the NAI reactivity of sgRNA M probed in infected Vero-E6 cells. Interestingly, the reactivity profile is similar for regions: 131–140, 160–172, 175–188, 219–228, 234–242, 245–252, 271–277, 279–291, 293–307, 342–357, 359–364, 366–374, 385–396, 401–406, 410–415, 419–426, 428–434, 438–445, 449–461, 494–499, 516–522, 525–530, 533–546, 559–566, 571–583, 585–601, 604–614, 630–637, 639–650, 680–692, 711–727, 733–743 (regions longer than five nucleotides were mentioned, nucleotides with reactivities 0–0.5, 0.5–0.7, and ≥ 0.7 were treated as unreactive, moderately reactive, and highly reactive positions, respectively). The reactivity of sgRNA M from living cells used in the RNAstructure program supported the existence of structural motifs in sgRNA M, such as 7–23, 131–150, 181–194, 231–252, 305–323, 340–363/434–458, 383–414, 517–590, 622–688, 712–758. The preservation of in vitro thermodynamically stable motifs in living cells could indicate that they are important in the viral replication cycle. On the other hand, some regions of sgRNA M from in vivo probing experiments have different reactivities than in vitro datasets, for example, 308–315, 322–332, 380–384. It is possible that these regions could interact with proteins or other cellular components.

We also compared long-range RNA–RNA interactions within the secondary structure of sgRNA M mapped in vitro (Figure 2) with the in vivo RNA–RNA interactome of sgRNA M [43]. Overall, these interactions are different and complex. Moreover, Ziv and coauthors discovered the co-existence of alternative SARS-CoV-2 gRNA and sgRNA topologies, held by long-range base-pairing between regions tens of thousands of nucleotides apart [43].

We additionally compared our sgRNA M model (Figure 2) with a proposed gene M secondary structure [20]. Generally, our presented model of the secondary structure of sgRNA M and the corresponding region of the whole SARS-CoV-2 genome obtained by probing in vivo [20] are different (Figure 3). This difference is in agreement with a previous study about the in vivo RNA–RNA interactome of the full-length SARS-CoV-2 genome and several sgRNAs. Here some structural aspects of viral RNA are also discussed [43]. This investigation revealed that the viral genome and subgenomes adopt alternative topologies inside the cell. Moreover, some long-range RNA–RNA interactions in sgRNA of SARS-CoV-2 are unique [43]. However, some regions in our presented model of sgRNA M fold in the same structure as in the genome M region of infected cells. Motifs: 153–167, 178–198, 223–259, 299–329, 340–350/448–458, 352–361/437–445, 365–433, 484–498, 514–592, 604–698, 712–759 are consistent with the in-cell secondary structure model of the corresponding region in the SARS-CoV-2 genome [20]. However, hairpin 514–592 is three base pairs longer and possesses an additional internal loop in the cellular model. In turn, hairpins 131–150 and 281–297 are almost identical to those of the corresponding region in the SARS-CoV-2 genome that are mapped in cells [20]. In our model, loops are longer than the corresponding region of the in-cell model. On the other hand, the small motif 503–512 is the only hairpin structure uniquely characteristic for sgRNA M and does not exist in the context of the SARS-CoV-2 genome. Our determined sgRNA M secondary structure is also similar to corresponding regions of other published whole-genome SARS-CoV-2 models [22,23]. This similarity between our sgRNA M structure and the in-cell determined structure of the M sequence in the whole genome context is surprising since that RNA structure in vitro and in vivo could be significantly different. In vivo interaction between RNA and proteins or other molecules can influence secondary structure [52]. These data

indicated that the sequence and thermodynamics alone are major determinants of sgRNA M structural motifs formation. All results point to these stable motifs being functionally significant. Furthermore, experiments and computational analyses have shown that large amounts of double-stranded regions have a strong propensity to interact with proteins and act as scaffolds for RNA-binding proteins [53–55]. sgRNA M is very structured, and it is possible for stable helices to interplay with proteins.

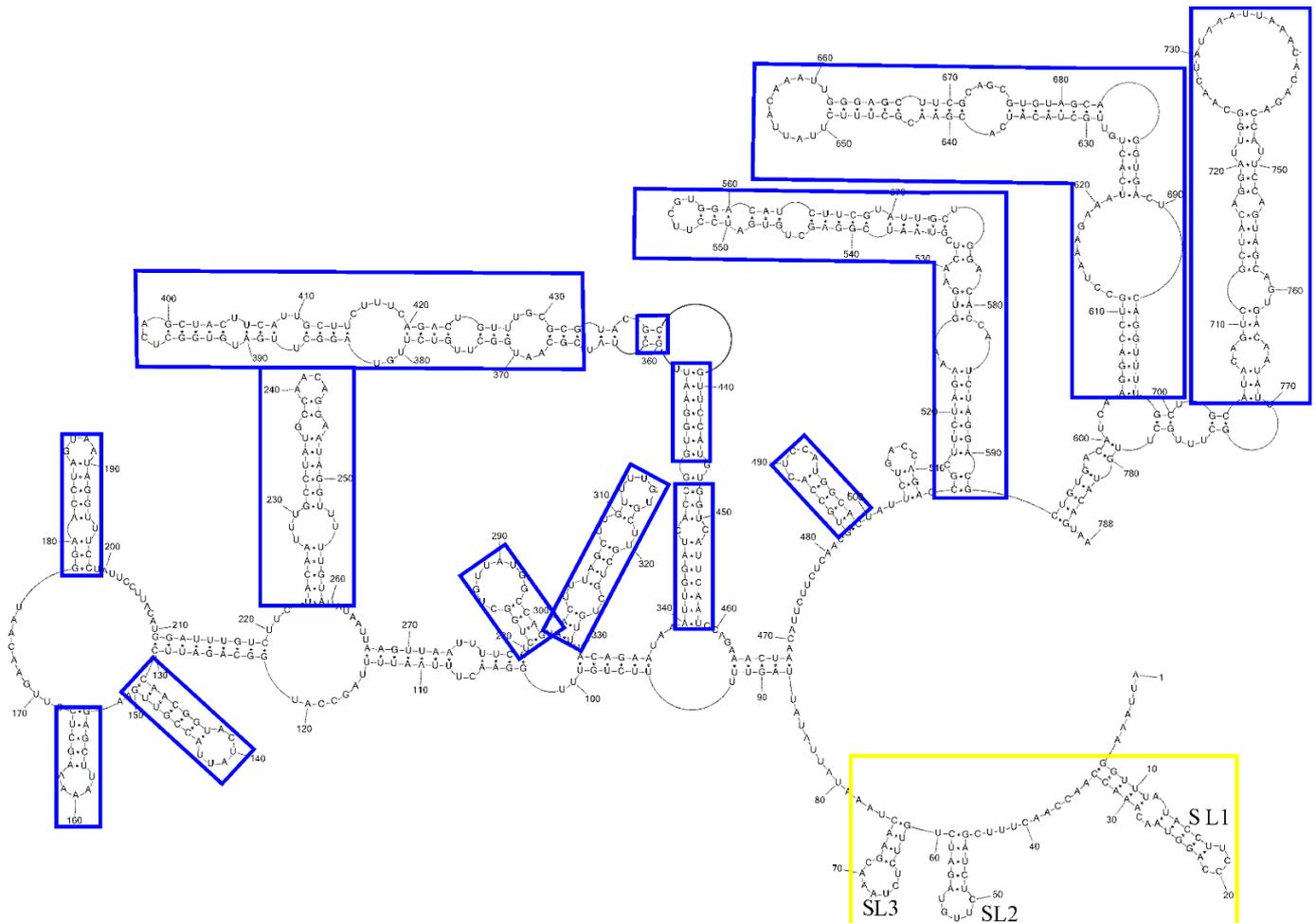


Figure 3. Comparison of the secondary structure of sgRNA M and its corresponding region of SARS-CoV-2 genome mapped in cells. Blue rectangle indicates the same base pairs within the sgRNA M model and the corresponding region of the SARS-CoV-2 genome mapped in cells [20]. Yellow rectangle indicates motifs of the leader sequence.

3.4. Local Structural Motifs in sgRNA M Are Mostly Independent of Leader Sequence

We compared our sgRNA M model with the corresponding region of an *in vitro* model of the whole SARS-CoV-2 genome [19] to check the influence of the 5' leader sequence on the folding of sgRNA. We indicated that the structures are different. This feature is consistent with a study of sgRNA N and its corresponding region in the SARS-CoV-2 genome. This data indicated that the same RNA sequences can fold in different structures in the subgenomic and genomic contexts [14]. However, some local motifs (Figure 2; 131–150, 153–167, 181–194, 223–259, 267–279/102–114, 280–297, 299–329, 331–335/101–105, 340–358/439–458, 372–426, 483–499, 514–592, 604–698, 712–758) within the sgRNA M and SARS-CoV-2 genome are identical. Motifs 604–698 is slightly different in both models. Similar motifs are stable independently of the neighboring regions (leader sequence in sgRNA M and sequences upstream/downstream of the M region in SARS-CoV-2 genome). It is possible that the existence and appropriate folding of neighboring regions has some,

but relatively small, influence on some local motifs of sgRNA M. Some local structure motifs are identical within in vitro and in vivo SARS-CoV-2 genome models [20].

3.5. Covariation Analysis of the sgRNA M Secondary Structure

The covariation analysis of the sgRNA M model presented here, utilizing the cm-builder pipeline with an alignment against 25,571 *Coronaviridae* sequences, yielded no base pairs with statistically significant covariation. Covariation (i.e., sites of mutated sequence which maintain base pairing and a 2D structure) is often used to support the potential for an RNA motif to be functional as the structure is being preserved even when the primary sequence is not. Importantly though, a lack of covariation does not indicate a lack of potential functionality.

An analysis of the conservation of the sgRNA M model base pairs against an alignment of 453 homologous sequences showed an average base pair conservation of 97.57%, with stem SL1 (Figure 2) having an average conservation of 50.97%, stem SL2 (Figure 2) averaging 83.48% and the remaining base pairs averaging roughly 100% conservation (Table S1, Supporting Information_3).

The high degree of conservation of most base pairs in the model may be partially responsible for the lack of detectable covariation. This lack of significant covariation is in line with previous studies [15,19,56]. Despite extensive evidence of stable, ordered secondary structure, few motifs were supported by significant covariation. Additionally, as most of the structures presented here exist within the sgRNA M coding sequence, there are additional evolutionary pressures to conserve these sequences as to not disrupt the protein amino acid sequence.

3.6. Possible Influence of Nucleotide Mutation on sgRNA M Structure of SARS-CoV-2 Variants

The SARS-CoV-2 genome constantly evolves, new mutations appear and virus variants are monitored. Changing of RNA sequence must occur in a frame to not be lethal for the virus and to preserve the function of proteins and also RNA structure. Emerging mutations should retain base pairs in RNA structure motifs that are important in the viral cycle. Therefore, we analyzed nucleotide mutations in the sequence of the M gene of SARS-CoV-2 variants. Table 3 display mutations based on SARS-CoV-2 variants and their representatives deposited in Global Initiative on Sharing Avian Influenza Data (GISAID). When compared to our model, most of the nucleotide mutations occur in single-stranded regions of the sgRNA M structure and have no influence on the base-pairing structure. Interestingly, three consistent mutations (preserving base pairs) are in the helix region of motif 94–130/211–336. Consistent mutations are additional support for the existence of these long-distance interactions. Two hairpin motifs in regions 153–167 and 299–329 (Figure 2) appear to have no functional significance because non-consistent mutations in these structures exclude their conservation (Table 3).

Table 3. Mutations in the sgRNA M coding region in SARS-CoV-2 variants.

Nucleotide Mutations in Coding Region ¹	SARS-CoV-2 Variants and Variants Representatives ²	Amino Acid Mutation ³	Influence of Nucleotide Mutations on RNA Structure ⁴
8(127) A > G	Omicron-B.1.1.529 (EPI ISL 6704867) Omicron BA.1 (EPI ISL 8482282) Omicron BA.2 (EPI ISL 8479001)	3D > G	consistent mutation A-U > G•U
46(165) C > A	AT.1 (EPI ISL 1652580)	16L > I	inconsistent mutation G-C > G..A
55(174) C > G	Omicron-B.1.1.529 (EPI ISL 6704867) Omicron BA.1 (EPI ISL 8482282) Omicron BA.2 (EPI ISL 8479001)	19Q > E	single-stranded region—no influence

Table 3. Cont.

Nucleotide Mutations in Coding Region ¹	SARS-CoV-2 Variants and Variants Representatives ²	Amino Acid Mutation ³	Influence of Nucleotide Mutations on RNA Structure ⁴
82(201) U > C	R.1 (EPI ISL 1123466)	28F > L	single-stranded region—no influence
84(203) C > U	B.1.619.1 (EPI ISL 2361101)	-	single-stranded region—no influence
85(204) C > U	C.1.2 (EPI ISL 2942287)	29L > F	single-stranded region—no influence
159(278) C > U	A.2.5.2 (EPI ISL 1502915) Kappa-B.1.617.1 (EPI ISL 1384866) Epsilon-B.1.429 (EPI ISL 648527) Epsilon-B.1.427 (EPI ISL 1531902)	-	consistent mutation C-G > U•G
187(306) G > A	Omicron-B.1.1.529 (EPI ISL 6704867) Omicron BA.1 (EPI ISL 8482282) Omicron BA.2 (EPI ISL 8479001) AV.1 B.1.1.482.1 (EPI ISL 2179526)	63A > T	inconsistent mutations G-C > A..C
213(332) C > U	B.1.466.2 (EPI ISL 1533080)	-	consistent mutation C-G > U•G
241(361) G > U	Delta-AY.27 B.1.617.2 (EPI ISL 3910943)	81A > S	single-stranded region—no influence
245(364) U > C	Delta-B.1.617.2 (EPI ISL 1758376) C.1.2 (EPI ISL 2942287) B.1.640 (EPI ISL 5655471) Eta-B.1.525 (EPI ISL 760883) C.36.3 (EPI ISL 1245879) B.1.1.318 (EPI ISL 986813) B.1.619.1 (EPI ISL 2361101) B.1.575 (EPI ISL 2634469) B.1.1.523 (EPI ISL 2448704) AZ.5 (EPI ISL 2834919) Delta-B.1.617.2 (EPI ISL 2519798) Delta-AY.4.2 B.1.617.2 (EPI ISL 2851674) Delta-AY.33 B.1.617.2 (EPI ISL 4506526) Delta-AY.27 B.1.617.2 (EPI ISL 3910943) Delta-AY.26 B.1.617.2 (EPI ISL 2306061) Delta-AY.25 (EPI ISL 2295285) Delta-AY.3 B.1.617.2 (EPI ISL 3459265) Delta-AY.43 (EPI ISL 3565601) Delta-AY.47 (EPI ISL 2611181) Delta-AY.98.1 (EPI ISL 3305991) Delta-B.1.617.2 (EPI ISL 2306894)	82I > T	single-stranded region—no influence
245(364) U > G	Kappa-B.1.617.1 (EPI ISL 1384866)	82I > S	single-stranded region—no influence
279(398) C > G	B.1.177.82 (EPI ISL 617709)	-	single-stranded region—no influence
372(491) C > U	Lambda-C.37 (EPI ISL 1138413) N10 (EPI ISL 1465243)	-	single-stranded region—no influence

Table 3. Cont.

Nucleotide Mutations in Coding Region ¹	SARS-CoV-2 Variants and Variants Representatives ²	Amino Acid Mutation ³	Influence of Nucleotide Mutations on RNA Structure ⁴
373(492) C > U	AV.1 B.1.1.482.1 (EPI ISL 2179526)	125H > Y	single-stranded region—no influence
450(569) U > C	B.1.258.17 (EPI ISL 618584)	-	single-stranded region—no influence
621(740) C > U	A.2.5.2 (EPI ISL 1502915)	-	single-stranded region—no influence

¹—The first number in the column indicates the nucleotide number from the 5' end of the M coding region; number in parenthesis means the corresponding nucleotide number of sgRNA M, as in Figure 2. ²—Variants and accession number of variant representatives in GISAID. ³—Number before mutation means amino acid number in M protein. ⁴—Consistent mutation means single point mutation that preserves pairing. “-” indicates Watson-Crick pair, “•” shows GU wobble pair, “..” indicates no Watson-Crick and GU interactions.

4. Conclusions

For the first time, the secondary structure of sgRNA M was determined based on the experimental data from several chemical mapping methods and bioinformatic analyses. The secondary structure model contains unique features likely to be important for sgRNA M functions. The structure also includes several of the same motifs as the genomic M fragment in the SARS-CoV-2 genome (Figure 3). Previously published reactivity of sgRNA M from structure probing in living cells supported the existence of some of the presented structural motifs of our sgRNA M model [23]. Although covariation analysis shows no base pairs with statistically significant covariation, the mutations of gene M within SARS-CoV-2 variants is largely in agreement with the presented structure and supports long-distance helices. This new knowledge about sgRNA M provides insights to better understand virus biology and could be used for anti-SARS-CoV-2 strategies and designing new therapeutics. The revealed unique or same as in gRNA M structural motifs could be promising targets for antisense oligonucleotides, siRNAs and small molecules

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/v14020322/s1>, Supporting_Information_1: Figure S1: Folding of sgRNA M on agarose gel; Figure S2: Example of capillary electrophoresis raw.data; Figure S3: sgRNA M nucleotides reactivity diagrams; Supporting_Information_2: Table S1: sgRNA M mapping data; Supporting_Information_3: Table S1: Base pairs counts for secondary structure of sgRNA M.

Author Contributions: Conceptualization, E.K. and M.S.-J.; methodology, data analysis, M.S.-J., A.R., W.N.M. and C.A.O.; investigation, M.S.-J. and R.K.; writing—original draft preparation, M.S.-J. and E.K.; writing—review and editing, E.K., R.K., A.R., W.N.M. and C.A.O.; visualization, M.S.-J.; supervision, E.K.; project administration, E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by National Science Centre grants UMO-2020/01/0/NZ6/00137 to E.K, UMO-2019/33/B/ST4/01422 to R.K and R01GM133810 to W.N.M.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available in the article and Supplementary Materials.

Conflicts of Interest: No potential conflict of interest was reported by the authors.

References

- Gloster, A.T.; Lamnisos, D.; Lubenko, J.; Presti, G.; Squatrito, V.; Constantinou, M.; Nicolaou, C.; Papacostas, S.; Aydın, G.; Chong, Y.Y.; et al. Impact of COVID-19 pandemic on mental health: An international study. *PLoS ONE* **2020**, *15*, e0244809. [CrossRef] [PubMed]

2. Rehman, S.U.; Shafique, L.; Ihsan, A.; Liu, Q. Evolutionary trajectory for the emergence of novel coronavirus SARS-CoV-2. *Pathogens* **2020**, *9*, 240. [[CrossRef](#)] [[PubMed](#)]
3. Kirtipal, N.; Bharadwaj, S.; Kang, S.G. From SARS to SARS-CoV-2, insights on structure, pathogenicity and immunity aspects of pandemic human coronaviruses. *Infect. Genet. Evol.* **2020**, *85*, 104502. [[CrossRef](#)] [[PubMed](#)]
4. Arya, R.; Kumari, S.; Pandey, B.; Mistry, H.; Bihani, S.C.; Das, A.; Prashar, V.; Gupta, G.D.; Panicker, L.; Kumar, M. Structural insights into SARS-CoV-2 proteins. *J. Mol. Biol.* **2021**, *433*, 166725. [[CrossRef](#)] [[PubMed](#)]
5. Wu, Y.C.; Chen, C.S.; Chan, Y.J. The outbreak of COVID-19: An overview. *J. Chinese Med. Assoc.* **2020**, *83*, 217–220. [[CrossRef](#)] [[PubMed](#)]
6. Romano, M.; Ruggiero, A.; Squeglia, F.; Maga, G.; Berisio, R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells* **2020**, *9*, 1267. [[CrossRef](#)] [[PubMed](#)]
7. V'kovski, P.; Kratzel, A.; Steiner, S.; Stalder, H.; Thiel, V. Coronavirus biology and replication: Implications for SARS-CoV-2. *Nat. Rev. Microbiol.* **2021**, *19*, 155–170. [[CrossRef](#)]
8. Kim, D.; Lee, J.Y.; Yang, J.S.; Kim, J.W.; Kim, V.N.; Chang, H. The Architecture of SARS-CoV-2 Transcriptome. *Cell* **2020**, *181*, 914–921. [[CrossRef](#)]
9. Hartenian, E.; Nandakumar, D.; Lari, A.; Ly, M.; Tucker, J.M.; Glaunsinger, B.A. The molecular virology of coronaviruses. *J. Biol. Chem.* **2020**, *295*, 12910–12934. [[CrossRef](#)]
10. Nomburg, J.; Meyerson, M.; DeCaprio, J.A. Pervasive generation of non-canonical subgenomic RNAs by SARS-CoV-2. *Genome Med.* **2020**, *12*, 108. [[CrossRef](#)]
11. Sola, I.; Almazán, F.; Zúñiga, S.; Enjuanes, L. Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu. Rev. Virol.* **2015**, *2*, 265–288. [[CrossRef](#)] [[PubMed](#)]
12. Thiel, V.; Ivanov, K.A.; Putics, Á.; Hertzog, T.; Schelle, B.; Bayer, S.; Weißbrich, B.; Snijder, E.J.; Rabenau, H.; Doerr, H.W.; et al. Mechanisms and enzymes involved in SARS coronavirus genome expression. *J. Gen. Virol.* **2003**, *84*, 2305–2315. [[CrossRef](#)] [[PubMed](#)]
13. Kiening, M.; Ochsenreiter, R.; Hellinger, H.J.; Rattei, T.; Hofacker, I.; Frishman, D. Conserved secondary structures in viral mRNAs. *Viruses* **2019**, *11*, 401. [[CrossRef](#)] [[PubMed](#)]
14. de Cesaris Araujo Tavares, R.; Mahadeshwar, G.; Wan, H.; Huston, N.C.; Pyle, A.M. The Global and Local Distribution of RNA Structure throughout the SARS-CoV-2 Genome. *J. Virol.* **2020**, *95*, e02190-20. [[CrossRef](#)]
15. Andrews, R.J.; O'Leary, C.A.; Tompkins, V.S.; Peterson, J.M.; Haniff, H.S.; Williams, C.; Disney, M.D.; Moss, W.N. A map of the SARS-CoV-2 RNA structure. *NAR Genomics Bioinforma.* **2021**, *3*, lqab043. [[CrossRef](#)]
16. Rangan, R.; Zheludev, I.N.; Hagey, R.J.; Pham, E.A.; Wayment-Steele, H.K.; Glenn, J.S.; Das, R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: A first look. *RNA* **2020**, *26*, 937–959. [[CrossRef](#)]
17. Simmonds, P. Pervasive RNA secondary structure in the genomes of SARS-CoV-2 and other coronaviruses. *MBio* **2020**, *11*, e01661-20. [[CrossRef](#)] [[PubMed](#)]
18. Vandelli, A.; Monti, M.; Milanetti, E.; Armaos, A.; Rupert, J.; Zacco, E.; Bechara, E.; Delli Ponti, R.; Tartaglia, G.G. Structural analysis of SARS-CoV-2 genome and predictions of the human interactome. *Nucleic Acids Res.* **2020**, *48*, 11270–11283. [[CrossRef](#)]
19. Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T.K.; Marinus, T.; Ogando, N.S.; Snijder, E.J.; van Hemert, M.J.; Bujnicki, J.M.; et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **2020**, *48*, 12436–12452. [[CrossRef](#)]
20. Lan, T.; Allan, M.; Malsick, L.; Khandwala, S.; Nyeo, S.; Sun, Y.; Guo, J.U.; Bathe, M.; Griffiths, A.; Rouskin, S. Insights into the secondary structural ensembles of the full SARS-CoV-2 RNA genome in infected cells. *bioRxiv* **2021**. 2020.06.29.
21. Sun, L.; Li, P.; Ju, X.; Rao, J.; Huang, W.; Ren, L.; Zhang, S.; Xiong, T.; Xu, K.; Zhou, X.; et al. In vivo structural characterization of the SARS-CoV-2 RNA genome identifies host proteins vulnerable to repurposed drugs. *Cell* **2021**, *184*, 1865–1883.e20. [[CrossRef](#)] [[PubMed](#)]
22. Sanders, W.; Fritch, E.J.; Madden, E.A.; Graham, R.L.; Vincent, H.A.; Heise, M.T.; Baric, R.S.; Moorman, N.J. Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. *bioRxiv* **2020**. 2020.06.15. [[CrossRef](#)]
23. Yang, S.L.; DeFalco, L.; Anderson, D.E.; Zhang, Y.; Aw, J.; Lim, S.Y.; Lim, X.N.; Tan, K.Y.; Zhang, T.; Chawla, T.; et al. Comprehensive mapping of SARS-CoV-2 interactions in vivo reveals functional virus-host interactions. *Nat Commun.* **2021**, *12*, 5113. [[CrossRef](#)] [[PubMed](#)]
24. Kierzek, E.; Kierzek, R. The thermodynamic stability of RNA duplexes and hairpins containing N6-alkyladenosines and 2-methylthio-N6-alkyladenosines. *Nucleic Acids Res.* **2003**, *31*, 4472–4480. [[CrossRef](#)]
25. Xia, T.; SantaLucia, J.; Burkard, M.E.; Kierzek, R.; Schroeder, S.J.; Jiao, X.; Cox, C.; Turner, D.H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson - Crick base pairs. *Biochemistry* **1998**, *37*, 14719–14735. [[CrossRef](#)]
26. Ehresmann, C.; Baudin, F.; Mougél, M.; Romby, P.; Ebel, J.P.; Ehresmann, B. Probing the structure of RNAs in solution. *Nucleic Acids Res.* **1987**, *15*, 9109–9128. [[CrossRef](#)]
27. Merino, E.J.; Wilkinson, K.A.; Coughlan, J.L.; Weeks, K.M. RNA structure analysis at single nucleotide resolution by Selective 2'-Hydroxyl Acylation and Primer Extension (SHAPE). *J. Am. Chem. Soc.* **2005**, *127*, 4223–4231. [[CrossRef](#)]
28. Ziehler, W.A.; Engelke, D.R. Probing RNA Structure with Chemical Reagents and Enzymes. *Curr. Protoc. Nucleic Acid Chem.* **2001**. Chapter 6, Unit 6.1. [[CrossRef](#)]

29. Karabiber, F.; McGinnis, J.L.; Favorov, O.V.; Weeks, K.M. QuShape: Rapid, accurate, and best-practices quantification of nucleic acid probing information, resolved by capillary electrophoresis. *RNA* **2013**, *19*, 63–73. [[CrossRef](#)]
30. Reuter, J.S.; Mathews, D.H. RNAstructure: Software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **2010**, *11*, 129. [[CrossRef](#)]
31. Deigan, K.E.; Li, T.W.; Mathews, D.H.; Weeks, K.M. Accurate SHAPE-directed RNA structure determination. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 97–102. [[CrossRef](#)] [[PubMed](#)]
32. Mathews, D.H.; Disney, M.D.; Childs, J.L.; Schroeder, S.J.; Zuker, M.; Turner, D.H. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of RNA secondary structure. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 7287–7292. [[CrossRef](#)] [[PubMed](#)]
33. Nawrocki, E.P.; Eddy, S.R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013**, *29*, 2933–2935. [[CrossRef](#)] [[PubMed](#)]
34. Nawrocki, E.P.; Kolbe, D.L.; Eddy, S.R. Infernal 1.0: Inference of RNA alignments. *Bioinformatics* **2009**, *25*, 1335–1337. [[CrossRef](#)]
35. Rivas, E.; Clements, J.; Eddy, S.R. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat. Methods* **2016**, *14*, 45–48. [[CrossRef](#)]
36. Rivas, E.; Clements, J.; Eddy, S.R. Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* **2020**, *36*, 3072–3076. [[CrossRef](#)]
37. Pickett, B.E.; Sadat, E.L.; Zhang, Y.; Noronha, J.M.; Squires, R.B.; Hunt, V.; Liu, M.; Kumar, S.; Zaremba, S.; Gu, Z.; et al. ViPR: An open bioinformatics database and analysis resource for virology research. *Nucleic Acids Res.* **2012**, *40*, D593–D598. [[CrossRef](#)]
38. Pickett, B.E.; Greer, D.S.; Zhang, Y.; Stewart, L.; Zhou, L.; Sun, G.; Gu, Z.; Kumar, S.; Zaremba, S.; Larsen, C.N.; et al. Virus pathogen Database and Analysis Resource (ViPR): A comprehensive bioinformatics Database and Analysis Resource for the Coronavirus research community. *Viruses* **2012**, *4*, 3209–3226. [[CrossRef](#)]
39. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2018**, *20*, 1160–1166. [[CrossRef](#)]
40. Lu, Z.J.; Gloor, J.W.; Mathews, D.H. Improved RNA secondary structure prediction by maximizing expected pair accuracy. *RNA* **2009**, *15*, 1805–1813. [[CrossRef](#)]
41. Miao, Z.; Tidu, A.; Eriani, G.; Martin, F. Secondary structure of the SARS-CoV-2 5'-UTR. *RNA Biol.* **2021**, *18*, 447–456. [[CrossRef](#)] [[PubMed](#)]
42. Wacker, A.; Weigand, J.E.; Akabayov, S.R.; Altincekic, N.; Bains, J.K.; Banijamali, E.; Binas, O.; Castillo-Martinez, J.; Cetiner, E.; Ceylan, B.; et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **2020**, *48*, 7204–7205. [[CrossRef](#)] [[PubMed](#)]
43. Ziv, O.; Price, J.; Shalamova, L.; Kamenova, T.; Goodfellow, I.; Weber, F.; Miska, E.A. The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell* **2020**, *80*, 1067–1077. [[CrossRef](#)] [[PubMed](#)]
44. Chen, S.C.; Olsthoorn, R.C.L. Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. *Virology* **2010**, *401*, 29–41. [[CrossRef](#)] [[PubMed](#)]
45. Li, L.; Kang, H.; Liu, P.; Makkinje, N.; Williamson, S.T.; Leibowitz, J.L.; Giedroc, D.P. Structural Lability in Stem-Loop 1 Drives a 5' UTR-3' UTR Interaction in Coronavirus Replication. *J. Mol. Biol.* **2008**, *377*, 790–803. [[CrossRef](#)] [[PubMed](#)]
46. Liu, P.; Li, L.; Millership, J.J.; Kang, H.; Leibowitz, J.L.; Giedroc, D.P. A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *RNA* **2007**, *13*, 763–780. [[CrossRef](#)]
47. Xu, L.; Khadijah, S.; Fang, S.; Wang, L.; Tay, F.P.L.; Liu, D.X. The Cellular RNA Helicase DDX1 Interacts with Coronavirus Nonstructural Protein 14 and Enhances Viral Replication. *J. Virol.* **2010**, *84*, 8571–8583. [[CrossRef](#)]
48. Vlachogiannis, N.I.; Verrou, K.-M.; Stellos, K.; Sfrikakis, P.P.; Paraskevis, D. The role of A-to-I RNA editing in infections by RNA viruses: Possible implications for SARS-CoV-2 infection. *Clin. Immunol.* **2021**, *226*, 108699. [[CrossRef](#)]
49. Sarkar, S.N.; Bandyopadhyay, S.; Ghosh, A.; Sen, G.C. Enzymatic characteristics of recombinant medium isozyme of 2'-5' oligoadenylate synthetase. *J. Biol. Chem.* **1999**, *274*, 1848–1855. [[CrossRef](#)]
50. Sarkar, S.N.; Ghosh, A.; Wang, H.W.; Sung, S.S.; Sen, G.C. The nature of the catalytic domain of 2'-5'-oligoadenylate synthetases. *J. Biol. Chem.* **1999**, *274*, 25535–25542. [[CrossRef](#)]
51. Wu, C.H.; Chen, P.J.; Yeh, S.H. Nucleocapsid phosphorylation and RNA helicase DDX1 recruitment enables coronavirus transition from discontinuous to continuous transcription. *Cell Host Microbe* **2014**, *16*, 462–472. [[CrossRef](#)] [[PubMed](#)]
52. Ponti, R.D.; Armaos, A.; Vandelli, A.; Tartaglia, G.G. CROSSalive: A web server for predicting the in vivo structure of RNA molecules. *Bioinformatics* **2020**, *36*, 940–941. [[CrossRef](#)] [[PubMed](#)]
53. de Groot, N.S.; Armaos, A.; Graña-Montes, R.; Alriquet, M.; Calloni, G.; Vabulas, R.M.; Tartaglia, G.G. RNA structure drives interaction with proteins. *Nat. Commun.* **2019**, *10*, 3246. [[CrossRef](#)] [[PubMed](#)]
54. Cid-Samper, F.; Gelabert-Baldrich, M.; Lang, B.; Lorenzo-Gotor, N.; Ponti, R.D.; Severijnen, L.A.W.F.M.; Bolognesi, B.; Gelpi, E.; Hukema, R.K.; Botta-Orfila, T.; et al. An Integrative Study of Protein-RNA Condensates Identifies Scaffolding RNAs and Reveals Players in Fragile X-Associated Tremor/Ataxia Syndrome. *Cell Rep.* **2018**, *25*, 3422–3434. [[CrossRef](#)]
55. Cerase, A.; Armaos, A.; Neumayer, C.; Avner, P.; Guttman, M.; Tartaglia, G.G. Phase separation drives X-chromosome inactivation: A hypothesis. *Nat. Struct. Mol. Biol.* **2019**, *26*, 331–334. [[CrossRef](#)]
56. Huston, N.C.; Wan, H.; Strine, M.S.; de Cesaris Araujo Tavares, R.; Wilen, C.B.; Pyle, A.M. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **2021**, *81*, 584–598.e5. [[CrossRef](#)]