RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses

Ramya Rangan[1], Ivan N. Zheludev[2], Rhiju Das[1,2,3*]

[1]Biophysics Program, Stanford University, Stanford CA 94305
[2]Department of Biochemistry, Stanford University School of Medicine, Stanford CA 94305
[3]Department of Physics, Stanford University, Stanford CA 94305

*Corresponding author: rhiju@stanford.edu

Abstract

As the COVID-19 outbreak spreads, there is a growing need for a compilation of conserved RNA genome regions in the SARS-CoV-2 virus along with their structural propensities to guide development of antivirals and diagnostics. Using sequence alignments spanning a range of betacoronaviruses, we rank genomic regions by RNA sequence conservation, identifying 79 regions of length at least 15 nucleotides as exactly conserved over SARS-related complete genome sequences available near the beginning of the COVID-19 outbreak. We then confirm the conservation of the majority of these genome regions across 739 SARS-CoV-2 sequences reported to date from the current COVID-19 outbreak, and we present a curated list of 30 'SARS-related-conserved' regions. We find that known RNA structured elements curated as Rfam families and in prior literature are enriched in these conserved genome regions, and we predict additional conserved, stable secondary structures across the viral genome. We provide 106 'SARS-CoV-2-conserved-structured' regions as potential targets for antivirals that bind to structured RNA. We further provide detailed secondary structure models for the 5´ UTR, frame-shifting element, and 3´ UTR. Last, we predict regions of the SARS-CoV-2 viral genome have low propensity for RNA secondary structure and are conserved within SARS-CoV-2 strains. These 59 'SARS-CoV-2-conserved-unstructured' genomic regions may be most easily targeted in primer-based diagnostic and oligonucleotide-based therapeutic strategies.

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) has caused a rapidly expanding global pandemic, with the COVID-19 outbreak responsible at this time for over 600,000 cases and 25,000 deaths. The emergence of this pandemic has revealed an urgent need for diagnostic and antiviral strategies targeting SARS-CoV-2. Like other coronaviruses, SARS-CoV-2 is a positive sense RNA virus, with a large RNA genome approaching nearly 30 kilobases in length. Its RNA genome contains protein-coding open reading frames (ORFs) for the viral replication machinery, structural proteins, and accessory proteins. The genome additionally harbors various *cis*-acting RNA elements, with structures in the 5´ and 3´ untranslated region (UTRs) guiding viral replication, RNA synthesis and viral packaging.[1] Conserved RNA elements offer compelling targets for diagnostics. In addition, such RNA elements may be useful targets for antivirals, a concept supported by the recent development of antisense oligonucleotide therapeutics and small-molecule RNA-targeting drugs for a variety of targets across infectious and chronic diseases.[2-4]

Conserved structured RNA regions have already been shown to play critical functional roles in the life cycles of coronaviruses. Most coronavirus 5´ UTR's harbor at least four stem loops, with many showing heightened sequence conservation across betacoronaviruses, and various stems demonstrating functional roles in viral replication.[5] Furthermore, RNA secondary structure in the 5´ UTR exposes a critical sequence motif, the transcriptional regulatory sequence (TRS), that forms long-range RNA interactions necessary for facilitating the discontinuous transcription characteristic to coronaviruses.[6] Beyond the 5´ UTR, the frame-shifting element (FSE) in the first protein-coding ORF (ORF1ab) includes a pseudoknot structure that is necessary for the production of ORF1a and ORF1b from two overlapping reading frames via programmed ribosomal -1 frame-shifting.[7] In the 3´ UTR, mutually exclusive RNA structures including the 3´ UTR pseudoknot control various stages of the RNA synthesis pathway.[8]

Beyond these canonical structured regions, the RNA structure of the SARS-CoV-2 genome remains mostly unexplored. Unbiased discovery of other conserved regions and/or structured regions in the virus has the potential to uncover further functional *cis*-acting RNA elements. Here, we analyze RNA sequence conservation across SARS-related betacoronaviruses and currently available SARS-CoV-2 sequences, and we identify structured and unstructured regions that are conserved in each sequence set; these intervals can provide starting points for a variety of diagnostic and antiviral development strategies (Fig. 1). To identify structured regions, we predict maximum expected accuracy structures around conserved regions and report the support of these single structures from predictions of each RNA's structural ensemble. We additionally identify thermodynamically stable secondary structures across the whole genome, finding that currently known structures fall within these predictions, but also identifying various new candidate structured regions. We pinpoint unstructured genome intervals by identifying bases with low average base-pairing probabilities. Finally, we present secondary structure models for key RNA structural elements of SARS-CoV-2 annotated in the betacoronavirus family.

Results

RNA sequence conservation in SARS-related betacoronaviruses and SARS-CoV-2

To identify potential regions of conserved RNA secondary structure in the virus, we located stretches of the SARS-CoV-2 genome with high RNA sequence conservation across SARS-related betacoronavirus full genome sequences. By identifying regions with high RNA sequence conservation as a first step, we reasoned that we would be more likely to filter for functionally relevant structures that must be conserved through virus evolution and thereby discover targets that are potentially less likely to develop resistance against therapeutics or to escape diagnosis as the

virus evolves. To ensure reasonable numbers of sequences while still focusing on conservation and structure patterns most relevant to the current pandemic, we chose to analyze not all betacoronaviruses but a subgroup of SARS-related betacoronaviruses. These include SARS, SARS-CoV-2, and SARS-related bat coronaviruses, but not MERS, MHV, or other betacoronaviruses which have been classified into distinct subgroups based on different sequence and structure features in, for example, their 5´ UTR's.[9]

We carried out this analysis beginning with three different sequence alignments. Each captures a range of complete genome sequences across the SARS-related betacoronviruses, but differ in the total number of sequences and in the redundancy of those sequences, as follows:

1. The first multiple sequence alignment (**SARSr-MSA-1**) was computed by aligning sequences curated by Ceraolo and Giorgi,[10] filtered by including only the reference genome sequence NC_0405512.2[11] from the SARS-CoV-2 sequence set, removing the two MERS sequences, and leaving in all remaining betacoronavirus whole genome sequences. This alignment captures a range of SARS-related bat coronavirus and SARS sequences with only 11 sequences. These sequences correspond well to the SARS-related group defined in Gorbalenya, Baker, *et al*.[12]
2. The second MSA (**SARSr-MSA-2**) was obtained from BLAST by searching for the 100 complete genome sequences closest to the SARS-CoV-2 reference genome. This alignment captures a larger set of SARS-CoV-2, SARS, and bat coronavirus sequences than SARSr-MSA-1 but includes many sequences with high pairwise similarity.
3. The final MSA (**SARSr-MSA-3**) was obtained by locating all complete genome betacoronavirus sequences from the NCBI database, and removing mutually similar sequences with a 99% sequence conservation cutoff. With 180 sequences with at most 99% pairwise sequence similarity, this MSA captures a broader set of betacoronaviruses than SARSr-MSA-1 and SARSr-MSA-2 but is more challenging to align due to higher sequence diversity.

We computed conserved regions as contiguous stretches of 15 nucleotides or longer that were 100% conserved (cutoff for SARSr-MSA-1), 98% conserved (cutoff for SARSr-MSA-2), or 54% conserved (cutoff for SARSr-MSA-3). Searching for conserved regions of 15 nucleotides or more enables the design of antisense oligonucleotides that fall within these stretches. The sequence conservation cutoffs chosen ensured that at least 75 candidate conserved stretches were used for further structure analysis for each MSA. When calculating sequence conservation at the 5´ and 3´ sequence ends of the sequence, we did not include sequences that included only leading or trailing sequence deletions up to that point to avoid sequencing artefacts.

In Fig. 2, we depict conserved regions (100% conservation cutoff, SARSr-MSA-1) alongside the genome coordinates for the reference SARS-CoV-2 sequence. We observe intervals of conservation in the 5´ UTR and 3´ UTR genome regions, as expected based on prior work demonstrating sequence conservation surrounding structured RNA elements in these regions,[13] but we also noted stretches of RNA sequence conservation within some viral ORFs.

Interestingly, in SARSr-MSA-1 and SARSr-MSA-2 we found that conserved stretches overlapped with previously curated Rfam[14] families for *Coronaviridae* RNA secondary structures: the frameshifting stimulation element (Rfam family RF00507), the 3´ UTR pseudoknot (Rfam family RF00165), and the 3´ stem-loop II-like motif (Rfam family RF00164) (Fig. 2). Locations for the frameshifting stimulation element, 3´ UTR pseudoknot, and 3´ stem-loop II-like motif were confirmed using Infernal,[15] with all regions discovered at an $E<10^{-4}$ threshold. We also found overlap between conserved stretches and additional 5´ UTR structures that have been established for previous coronaviruses, including the original SARS virus, including stem loops 2-3 (SL2-3) and stem loop 5 (SL5).[16] These five known RNA structures overlap with conserved regions more than expected; in

10,000 random trials, the chance that five randomly chosen intervals of these lengths all overlap with the conserved regions from SARSr-MSA-1 or SARSr-MSA-2 is less than 0.0003. The enrichment of known RNA structures in these conserved regions suggests that other conserved regions may also harbor RNA structures.

To further tighten this list of conserved sequences to ones most relevant to the current COVID-19 outbreak, we analyzed whether sequence regions conserved across SARS and bat coronaviruses remain conserved in the SARS-Cov-2 strains, most of which emerged after our analysis above (Fig. 1A). We determined the conservation of conserved genome regions from SARSr-MSA-1 across SARS-CoV-2 sequences as of deposition date 03-18-20. For this analysis, we obtained two whole-genome multiple sequence alignments, keeping only full-length genome sequences of at least 29,000 nucleotides in both cases: the first includes 103 NCBI sequences (**SARS-CoV-2-MSA-1**), and the second includes 739 sequences deposited to GISAID[17] (**SARS-CoV-2-MSA-2**). We noted conserved regions in the betacoronavirus alignment SARSr-MSA-1 were more likely to be at least 99% conserved in both SARS-CoV-2-MSA-1 and SARS-CoV-2-MSA-2 than random intervals of the same size (binomial test p-value < 1e-5). Table 1 lists these regions, which we term the **SARS-related-conserved** regions. These genome regions are conserved across the betacoronavirus sequences in SARSr-MSA-1 and have at least 99% sequence conservation across whole-genome sequences from the SARS-CoV-2 outbreak as of March 18, 2020 (SARS-CoV-2-MSA-2).

Conservation percentages for SARSr-MSA-1, SARSr-MSA-2, SARS-CoV-2-MSA-1, and SARS-CoV-2-MSA-2 are included in Supplementary File 1. We expect that some diagnostic and therapeutic strategies will benefit from focusing on conserved regions across a broad range of betacoronaviruses, whereas others may benefit from focusing on regions conserved only in SARS-CoV-2; we revisit the latter category of SARS-CoV-2-unique regions below.

Predictions for structured regions in SARS-CoV-2

The intrinsic RNA structure of a conserved genome region is of interest in current medical research (Fig. 1B). On one hand, stable secondary structure domains are candidates for harboring stereotyped 3D RNA folds that present targets for small-molecule drug therapeutics. On the other hand, if an RNA region is sufficiently unstructured to allow binding by hybridization probes, antisense oligonucleotides may be used to disrupt these functional structures. Such unstructured stretches may also be more likely to be accessible to diagnostic and antiviral interventions including standard RT-PCR assays.

We used two approaches to make predictions for conserved structured regions in SARS-CoV-2. First, we predicted RNA structures centered on the most sequence-conserved regions of SARS-related betacoronavirus genomes (alignment SARSr-MSA-1). For each conserved stretch (at least 15 nucleotides long, 100% sequence conservation) along with 20 nucleotide flanking windows, we predicted maximum expected accuracy (MEA) secondary structures using Contrafold 2.0.[18] We then sought to rank sequences based on the predicted probability that the RNA folds into the MEA structure and not other structures. For this ranking, we used the estimated Matthews correlation coefficient (MCC) from each construct's base-pairing probability matrix.[19] We note here that while MCC is often used in the RNA structure modeling literature to assess agreement of a prediction with a reference structure, we here use the metric to assess how tightly concentrated the ensemble of predicted secondary structures is to a single predicted secondary structure, the MEA structure. An MEA structure with a higher estimated MCC is expected to have unpaired and paired bases that better align with the construct's predicted ensemble base-pairing probabilities, lending support to the single-structure MEA prediction. In Fig. 2 regions A-E, we display the five conserved regions with the top maximum expected accuracy (MEA) secondary structures as ranked by the estimated MCC (all regions listed in Supplementary File 1). Regions D and E occurred within the 5´ UTR and correspond to known SARS-related virus stem loops SL5a and SL2, respectively. Interestingly, region A is close

to but does not overlap with the frameshifting stimulation element; it lies 200 nucleotides downstream of the FSE and could perhaps be involved in a more elaborate structure, as has been described for human coronavirus 229E and other coronaviruses.[20]

We also sought independent methods to identify thermodynamically stable and conserved RNA structures, without initially guiding the search to focus on extremely sequence-conserved genome regions. We made predictions for structured regions using RNAz[21], beginning with the betacoronavirus alignment SARSr-MSA-1. RNAz predicts structured regions that are more thermodynamically stable than expected by comparison to random sequences of the same length and sequence composition (z-score), and additionally assesses regions by the support of compensatory and consistent mutations in the sequence alignment (SCI score). These two criteria are combined into a single P-score, which when tested empirically on a set of ncRNAs produced a false-positive rate of 4% at a P>0.5 cutoff and 1% at a P>0.9 cutoff. To predict structured regions across the full viral genome, we scanned the SARSr-MSA-1 alignment in windows of length 120 nucleotides sliding by 40 nucleotides, predicted all RNAz hits in the plus strand at a P>0.5 cutoff, clustered the resulting hits to generate maximally contiguous loci of the genome with predicted structure, and filtered results to only include loci with at least one window with a P>0.9 structure prediction.

The RNAz approach led to the prediction of 44 structured genome loci comprising 117 windows with predicted structure (P>0.9), with these loci covering 46% of the SARS-CoV-2 genome (Figure 3). We found that five canonical RNA structures (the frameshifting element, the 3´ UTR pseudoknot, the 3´ UTR hypervariable region, 5´ UTR SL2-3, and 5´ UTR SL5) were present in these loci. Additionally, conserved SARS-CoV-2 regions overlap significantly with predicted RNAz loci, with 62 of 78 SARS-CoV-2 conserved intervals at a 97% sequence cutoff overlapping by at least 15 nucleotides with RNAz loci. This enrichment is statistically significant (p-value<0.001 from comparisons to 10,000 random placements of conserved intervals). This enrichment also holds when considering overlaps with conserved regions from SARSr-MSA-1; 124 of the 229 SARSr-MSA-1 conserved intervals at a 90% conservation cutoff overlap by at least 15 nucleotides with RNAz loci (p-value 0.0038). This analysis potentially expands the set of conserved structural regions of SARS-CoV-2 beyond known Rfam families and those noted in the literature (full set of RNAz loci in Supplementary File 1). Top-scoring structured windows from RNAz that overlap with conserved sequence regions in SARS-CoV-2-MSA-2 for at least 15 nucleotides are included in Table 2; we termed these **SARS-CoV-2-conserved-structured** regions. Overlapping intervals between the RNAz predictions and conserved sequence regions in SARSr-MSA-1 are included in Supplementary File 1.

We sought to further check structured windows reported by RNAz using orthogonal approaches. First, we explored using R-scape to make structure predictions with covariation signal in the sequence alignments. However, we found that the SARSr-MSA-1 alignment had insufficient variation to detect conserved base pairs with covariation, lacking alignment power for all genomic windows.[22] Second, we validated structured window predictions with alifoldz, a program that calculates a z-score for an alignment window by comparing the window's consensus minimum free energy structure to that of random shuffled alignments. To mirror the RNAz analysis above, we scanned through windows of length 120 nucleotides sliding by 40 nucleotides. We chose a z-score cutoff of -2.69, which kept only 1% of windows when running alifoldz on all shuffled windows across the genome. This approach led to predicting 228 alifoldz structured windows, overlapping with 104 of the 117 RNAz structured windows (P>0.9 cutoff). This overlap is statistically significant (p-value<1e-05). RNAz structured windows supported by alifoldz analysis are highlighted in Supplementary File 1.

Conserved unstructured regions of SARS-CoV-2

We additionally located conserved regions of the viral genome predicted to *lack* structure, as such regions may be desired targets for some diagnostic and therapeutic approaches (Fig. 1C). We

scanned the SARS-CoV-2 reference genome in windows of length 120 nucleotides sliding by 40 nucleotides, and for each window, we predicted the base-pair probability matrix with Contrafold 2.0, using these probabilities to assemble average single-nucleotide base pairing probabilities across the genome. In Figure 3, we display the 76 stretches of the genome of length at least 15 nucleotides where every base has average base-pairing probability at most 0.4.

It is interesting to note that some structured 120 nucleotide windows reported by RNAz include these unpaired stretches. A potential explanation for this observation is that such regions encode for well-defined, conserved RNA structures that themselves harbor long unpaired loops to recruit proteins, distal RNA elements, or other molecular machinery.

Overall, we find that 58 of these unpaired stretches have at least 15 nucleotides of overlap with sequence regions that are at least 97% conserved in SARS-CoV-2-MSA-2 (Fig. 5). These unpaired stretches termed **SARS-CoV-2-conserved-unstructured** regions are listed in Table 3 (overlaps with SARSr-MSA-1 are included in Supplementary File 1.)

As an orthogonal check for the unstructured intervals predicted using Contrafold 2.0 base-pairing probabilities, we used Vienna's RNAplfold to compute unpaired probabilities for each genome position. In general, we found that RNAplfold predicted lower unpaired probabilities than Contrafold 2.0, with only 10 intervals of length at least 15 nucleotides having at least 0.6 probability of being unpaired, in contrast with the 76 stretches predicted by Contrafold 2.0. Nevertheless, we found that 9 of the 10 intervals predicted by Vienna's RNAplfold overlap with unpaired intervals predicted from our Contrafold 2.0 analysis (regions listed in Supplementary File 1.)

Secondary structure models for canonical structured regions of SARS-CoV-2

Currently known RNA structures that recur across betacoronaviruses provide potential starting points for therapeutic development targeting the SARS-CoV-2 RNA genome. Here, we include secondary structures for the 5´ UTR (Figure 4a), frame-shifting element (Figure 4b), and 3´ UTR (Figure 4c) for SARS-CoV-2, built by analyzing homology to literature-annotated structures in related betacoronaviruses. We additionally include computer-readable secondary structures in Table 4 and Supplementary File 1. A brief review of salient secondary structure features in these regions and their putative functional roles in the betacoronavirus life cycle follows.

The 5´ UTR includes five confident stem-loop structures (SL1-SL5), with structures verified by chemical mapping experiments in related coronaviruses.[9, 16] SL1 and SL2 are conserved across betacoronaviruses, with SL2 having the highest sequence conservation across the 5´ UTR.[13] The high A-U base-pairing content in the SARS-CoV-2 SL1 sequence and the bulged nucleotides align with prior reports that SL1 is relatively thermodynamically unstable to allow for the formation of long-range interactions.[23] SL2 has been shown to be critical for subgenomic RNA synthesis, with mutations in its conserved pentaloop retaining the production of genome-sized RNA, but not subgenomic RNA segments.[24] SL3, conserved only in betacoronaviruses, presents the transcription-regulating sequence (TRS) that base pairs with one of several complementary sequences in nascent negative-sense strands in a 'copy-choice mechanism' that gives rise to discontinuous transcription of subgenomic mRNAs.[6] SL4 contains a short upstream ORF, here labeled uORF, which precedes the first longer ORF1ab of the genome. The uORF leads to attenuated transcription of ORF1ab that appears helpful but is not essential for viral replication.[13] SL5 has been implicated in a potential role in viral packaging, and contains the AUG start codon for long ORF1ab which encodes the viral replicase/transcriptase polyprotein. The SARS-CoV-2 SL5 domain has common features with the domain in other group IIb betacoronaviruses, for instance including UUCGU pentaloops on SL5a and SL5b, and a GNRA tetraloop on SL5c.[9] Prior DMS-probing data for Stem 5 in SARS-CoV aligned with the proposed SL5a,b,c structures.[9] Two additional stems (SL6 and SL7) are predicted

from computer modeling here, but prior literature has not established whether such stems embedded in the coding region are functionally important across betacoronaviruses.

The frameshifting element (FSE) is located in ORF1ab and is involved in regulating a (-1) ribosomal frameshift event that is necessary for producing ORF1b. The FSE consists of a conserved pseudoknot structure that regulates the rate of ribosomal frameshifting at an upstream slippery site.[7] This domain is nearly exactly conserved between SARS-CoV and SARS-CoV-2, suggesting a similar mechanism for ribosomal pausing and slippage between the two viruses.[25]

The 3´ UTR contains various domains critical for regulating viral RNA synthesis and potentially translation. The most 5´ region of the 3´ UTR includes a switch-like domain involving mutually exclusive formation of a pseudoknot and stem-loop, both of which are essential for viral replication with putative roles in establishing the kinetics of RNA synthesis.[6, 26] The hyper-variable region (HVR) is not essential for viral RNA synthesis, as this can be removed while allowing for viral replication in tissue culture; however, viruses without this domain have lower pathogenicity in mice.[27] This domain contains a completely conserved octonucleotide sequence with unconfirmed functional significance. The stem-loop II-like motif (s2m) is another subregion of the HVR that is conserved in SARS-CoV-2 and other coronaviruses. A crystal structure of the SARS s2m domain has been shown to be homologous to an rRNA loop that binds translation initiation proteins, leading this domain to have a proposed role in recruiting host translational machinery.[28] The domain has been proposed to be a selfish element due to its recurrence in numerous virus families outside the *Coronaviridae*, but its function is not well understood.[29]

Discussion

Understanding the RNA structure of the SARS-CoV-2 genome can guide RNA-targeting interventions and diagnostics. Here we have presented an initial analysis of RNA sequence conservation across betacoronaviruses and current SARS-CoV-2 sequences, predictions for structured and unstructured domains of the viral RNA genome, and homology-derived secondary structure models for classic structured elements of the SARS-CoV-2 genome: the 5´ UTR, the frame-shifting element, and the 3´ UTR. By filtering for sequences that have more than one of these properties, we have curated three sets of RNA genomic regions of potential interest for further structural analysis, which we have termed the SARS-related-conserved, SARS-CoV-2-conserved-structured, and SARS-CoV-2-conserved-unstructured sets. Fig. 5 gives a more extensive presentation of how these sets overlap. Our hope is that these steps will provide useful starting points for efforts to develop antivirals and diagnostics that depend on targeting either structured or unstructured viral genomic regions.

The abundant RNA structures involved in the replication cycle of betacoronaviruses present ample opportunities for therapeutic development, but our analysis is not complete. First, while homology to prior structures annotated in betacoronaviruses lends some confidence to the 5´ UTR, frame-shifting element, and 3´ UTR secondary structures presented here, there may still be some inaccuracies. For instance, SL6 and SL7 in the 5´ UTR are built based on computer modeling. As another example, the frame-shifting element structure presented here differs in two base pairs compared to that presented by Kelly and Dinman.[25] Additional biochemical and genetic verification, particularly through compensatory mutagenesis, will be needed to further assess these structures.

Beyond the secondary structures highlighted here, prior work has pinpointed a variety of RNA-RNA interactions important to the betacoronavirus life cycle. Long-range interactions between the 5´ UTR and 3´ UTR have been implicated in RNA synthesis, for instance with mutations in the 5´ UTR SL1 only supporting viral replication when co-evolving with specific mutations in the 3´ UTR.[13, 30] Such long-range RNA-RNA interactions would be missed in our analyses above which have focused on

shorter windows. In related coronaviruses, RNA structures that act as packaging signals have been identified in genomic ORFs.[13] Such packaging signals may reside in regions identified here from our RNAz computational analysis or may have been missed.

We believe a more thorough structure/function analysis of the virus should be obtainable with recent experimental technologies that integrate multidimensional chemical mapping, electron microscopy, and computer modeling.[31-32] New candidates for structured RNA elements in SARS-CoV-2 could play various functional roles, perhaps regulating viral packaging, replication, RNA synthesis, or translation initiation. To further improve these structure predictions, information about protein-binding events should be integrated, and biochemical assays can be conducted with these proteins present or even in cells. Accounting for protein-binding events will be critical to completing a picture of accessible and structured RNA sites.

The secondary structures predicted here present a reasonable starting point for 3D modeling of RNA-only structures in various regions, including the 5´ UTR stem loop 5, the frame-shifting element, the 3´ UTR pseudoknot, and the 3´ stem-loop II-like motif. Furthermore, these regions and novel predicted structures can serve as candidates for RNA-only structure determination. Such 3D structures have the potential to reveal well-defined 3D folds with conserved binding domains for small-molecule drugs, potentially presenting alternative approaches for targeting SARS-CoV-2.

Methods

Conservation analysis

Three alignments for SARS-related viruses were prepared:

1. SARSr-MSA-1 was generated by re-aligning the sequences curated by Ceraolo and Giorgi[10] with MUSCLE[33] using default alignment settings, excluding all non-reference genome copies of the SARS-CoV-2 sequence and excluding MERS sequences JX869059.2 and KT368829.1.
2. SARSr-MSA-2 was generated by downloading the MSA provided by BLAST for the top 100 complete genome sequences closest to the SARS-CoV-2 reference genome.
3. SARSr-MSA-3 was generated by obtaining all complete betacoronavirus genome sequences available from the NCBI database, removing mutually similar sequences using a 99% cutoff with CD-HIT-EST[34], and computing an MSA with Clustal Omega[35] using default settings.

Two alignments for SARS-CoV-2 sequences were prepared:

1. SARS-CoV-2-MSA-1 was generated by downloading the MSA provided by NCBI for the 103 whole-genome SARS-CoV-2 sequences deposited as of 03-18-20.
2. SARS-CoV-2-MSA-2 was generated from 739 GISAID[17] sequences, including all sequences described in the Nextstrain project metadata file as of 03-18-20 (https://github.com/nextstrain/ncov). Sequences were aligned using MAFFT[36] with the --add flag to add GISAID sequences to seed alignment SARS-CoV-2-MSA-1, and the sequences in SARS-CoV-2-MSA-1 were subsequently removed to avoid duplicates.

All alignments are included in the associated GitHub repository: https://github.com/DasLab/SARSCoV2_Secstruct_Cons.

Analysis of structured elements

To identify regions in the SARS-CoV-2 reference genome NC_0405512.2[11] that were matches to Rfam[14] families, we used Infernal[15] to build covariance models from Rfam families RR0164, RR0165, and RR0507 with cmbuild, and we ran cmscan to find hits with an E<1e-04 threshold.

MEA structures were computed using Contrafold 2.0[18] for 20 nucleotide flanking windows around conserved intervals in SARS-CoV-2. To estimate the Matthews correlation coefficient of these single-structure predictions, we computed the pseudo-expected Matthews correlation coefficient as described in Hamada, et al.[19] Base-pairing probability matrices were computed Contrafold 2.0, and these were then used to calculated the expected number of true positive, true negative, false positive, and false negative base pairs. These computations were carried out using the Arnie package (https://github.com/DasLab/arnie).

RNAz[21] structures were predicted in windows of the SARS-CoV-2 genome using the SARSr-MSA-1 alignment. We used rnazWindow.pl to compile alignment windows across SARSr-MSA-1 with at least 4 sequences in each window, using a window size of 120 nucleotides sliding by 40 nucleotides, and using default settings otherwise. RNAz hits were computed at the P>0.5 threshold for the forward strand with z-scores computed without a shuffled sequence background for efficiency, using the --no-shuffle flag. The resulting RNAz structured windows were then clustered with rnazCluster.pl, filtered with rnazFilter.pl at a P>0.9 threshold, and sorted with rnazSort.pl.

We ran alifoldz[37] on the same genome windows used with RNAz above, again using SARSr-MSA-1. The alifoldz z-score computations were calculated for the forward strand only with alifoldz.pl. We additionally calculated alifoldz z-scores for alignment windows that were shuffled with shuffle-aln.pl to assess background z-scores, determining that 1% of shuffled alignment z-scores were less than -2.69.

We computed alignment powers with R-scape[38] to assess the potential for using SARSr-MSA-1 for covariation analysis. We generated Stockholm alignment files with biopython[39] for windows of 120 nucleotides each sliding by 40 nucleotides. For each window, we ran R-scape with the –fold flag to predict new structures, obtaining estimates for the power of each base pair in the predicted structure (here, power is the expected sensitivity for detecting base pairs given the number of substitutions in the alignment at that base pair). We then averaged across base pairs in each structure to obtain the alignment power as described in Rivas, et al.,[22] noting that all windows' alignment powers fell below the 0.10 threshold used by Rivas, et al.[22] to distinguish low-power from high-power alignments.

Analysis of unstructured elements

To obtain probabilities that each genome position in SARS-CoV-2 was unpaired, we computed base-pairing probability matrices with Contrafold 2.0[18] in windows of 120 nucleotides sliding by 40 nucleotides, and for each genome position, we summed the probabilities of pairing with all potential partners. We then averaged these nucleotide pairing probabilities across all windows that nucleotide was present in. Additionally, RNAplfold[40] was run with window size 120 nucleotides, producing another set of unpaired probabilities for each position in the genome.

Code and data availability

Code used for the conservation, structured, and unstructured analyses above can be found at the GitHub repository: https://github.com/DasLab/SARSCoV2_Secstruct_Cons. The repository additionally includes alignment files, Rfam families and covariance models, and output from the RNAz, R-scape, alifoldz and RNAplfold analyses.

Acknowledgments

The authors acknowledge support from the National Science Foundation Graduate Research Fellowship Program under grant no. 1650114 (R.R.), a Stanford Graduate Fellowship (I.N.Z.), and

NIH grant MIRA R35 GM122579 (to R.D.). We would like to thank Dr. Rachel H. Saluti, Dr. Edward A. Pham, and Dr. Jeffrey S. Glenn for providing advice on the conserved, structured intervals desirable for antisense oligonucleotide design; Hannah Wayment-Steele for useful discussions on secondary structure modeling; Andrew M. Watkins for updates to the RiboDraw software for rendering secondary structures; and Dr. Paul Gardner for providing rapid reviews of the initial preprint of this manuscript on bioRXiv.

Figures

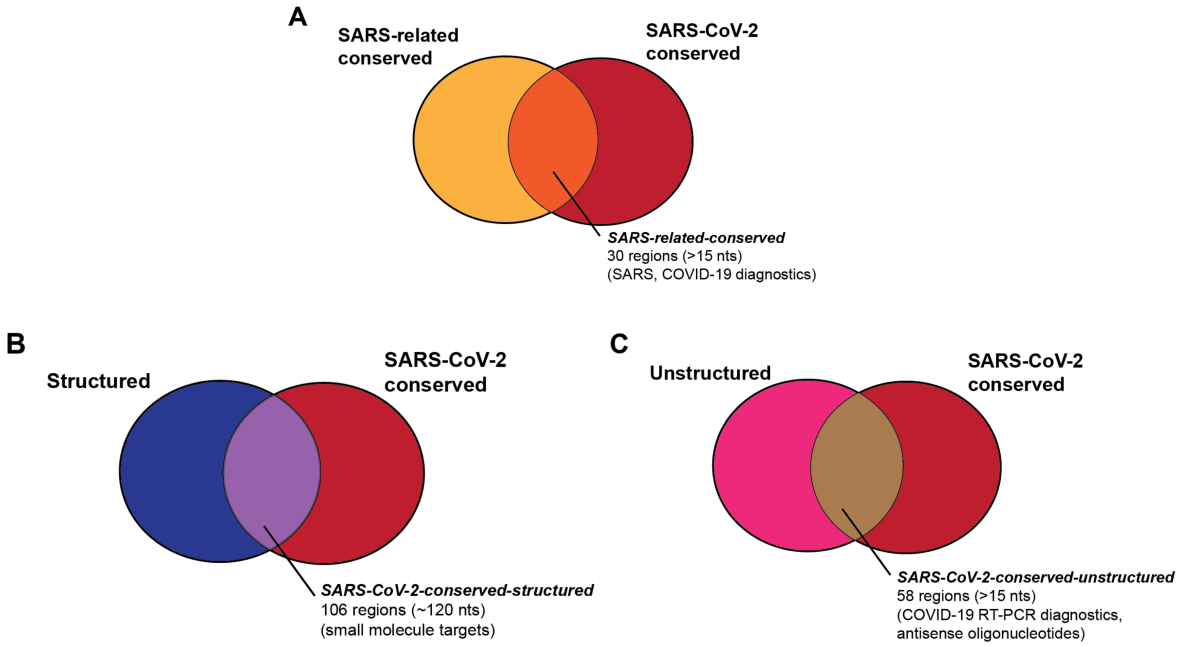

Figure 1: We aim to provide a series of genome regions in SARS-CoV-2 that are useful for a variety of diagnostic and therapeutic strategies, including regions that are (A) conserved in SARS-related betacoronaviruses and SARS-CoV-2 sequences (Table 1), (B) regions that are structured and conserved in SARS-CoV-2 sequences (Table 2), and (C) regions that are unstructured and conserved in SARS-CoV-2 sequences (Table 3).

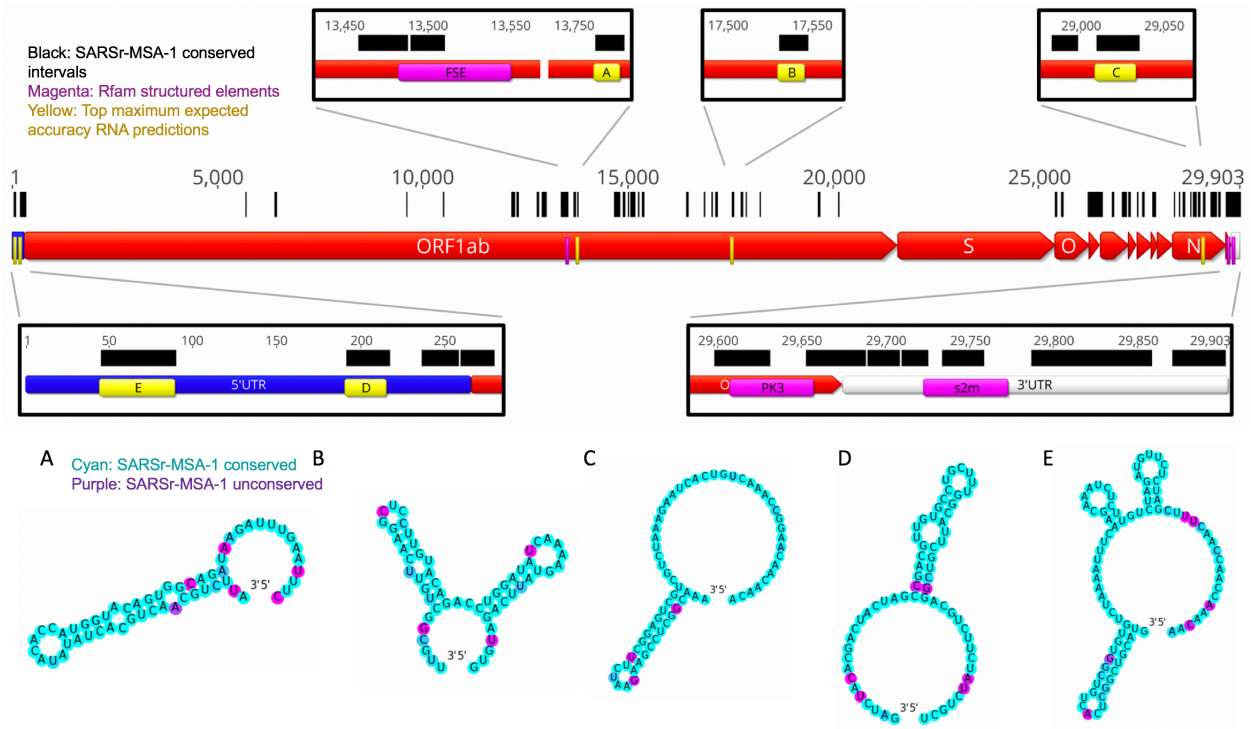

Figure 2: In black we annotate SARSr-MSA-1 conserved regions of the genome, superimposed on SARS-CoV-2 genome ORFs. We depict the top secondary structures as ranked by Matthews correlation coefficient that overlap with these conserved regions, ordered from A to E. Regions A to E are annotated on the genome in yellow and are located at genome positions: A:13743-13798, B:17511-17566, C:28990-29054, D:172-236, E:26-109. Secondary structures are colored by sequence conservation in SARSr-MSA-1 (cyan = more conserved, purple = less conserved). In magenta are depicted curated Rfam families present in coronaviruses, including the frame-shifting element (FSE), the 3´ UTR pseudoknot (PK3), and the 3´ stem-loop II-like motif (s2m). Figures prepared in Geneious[41] and draw_rna (https://github.com/DasLab/draw_rna).

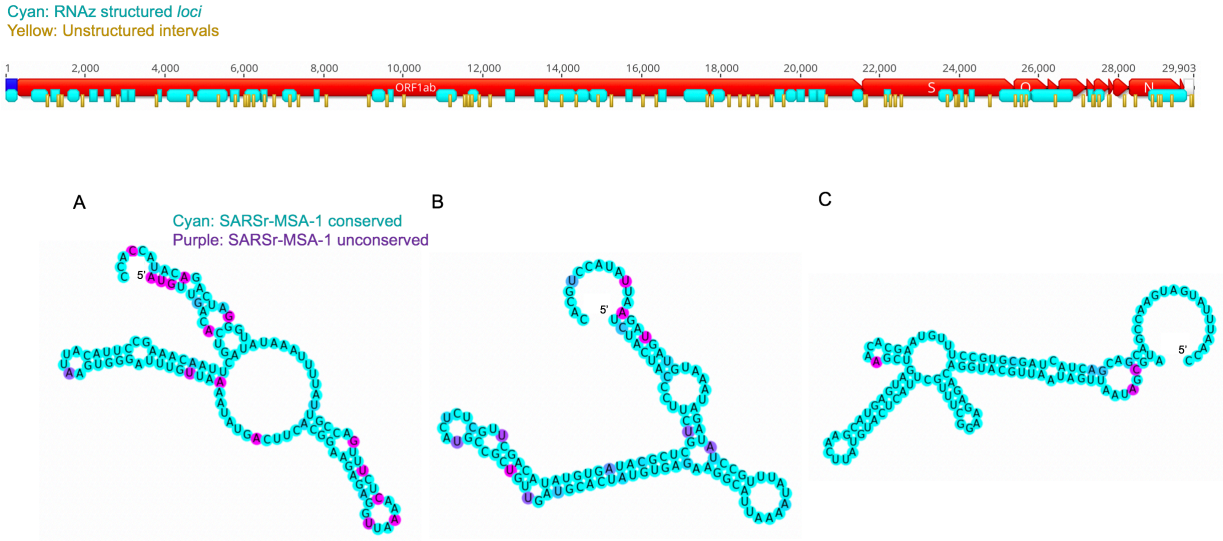

Figure 3: Structured (cyan) and unstructured (yellow) intervals on the genome ORFs for SARS-CoV-2, predicted from RNAz and a Contrafold 2.0 analysis, respectively. (A-C) highlight the three secondary structures for windows that do not overlap with known Rfam or literature-annotated structures with the highest P-value scores from RNAz (all P>0.9). These windows are located at genome positions 14207-14366 (A), 17126-17245 (B), and 26176-26295 (C). Secondary structures are colored by sequence conservation (cyan = more conserved, purple = less conserved). Figures prepared in Geneious[41] and draw_rna (https://github.com/DasLab/draw_rna).

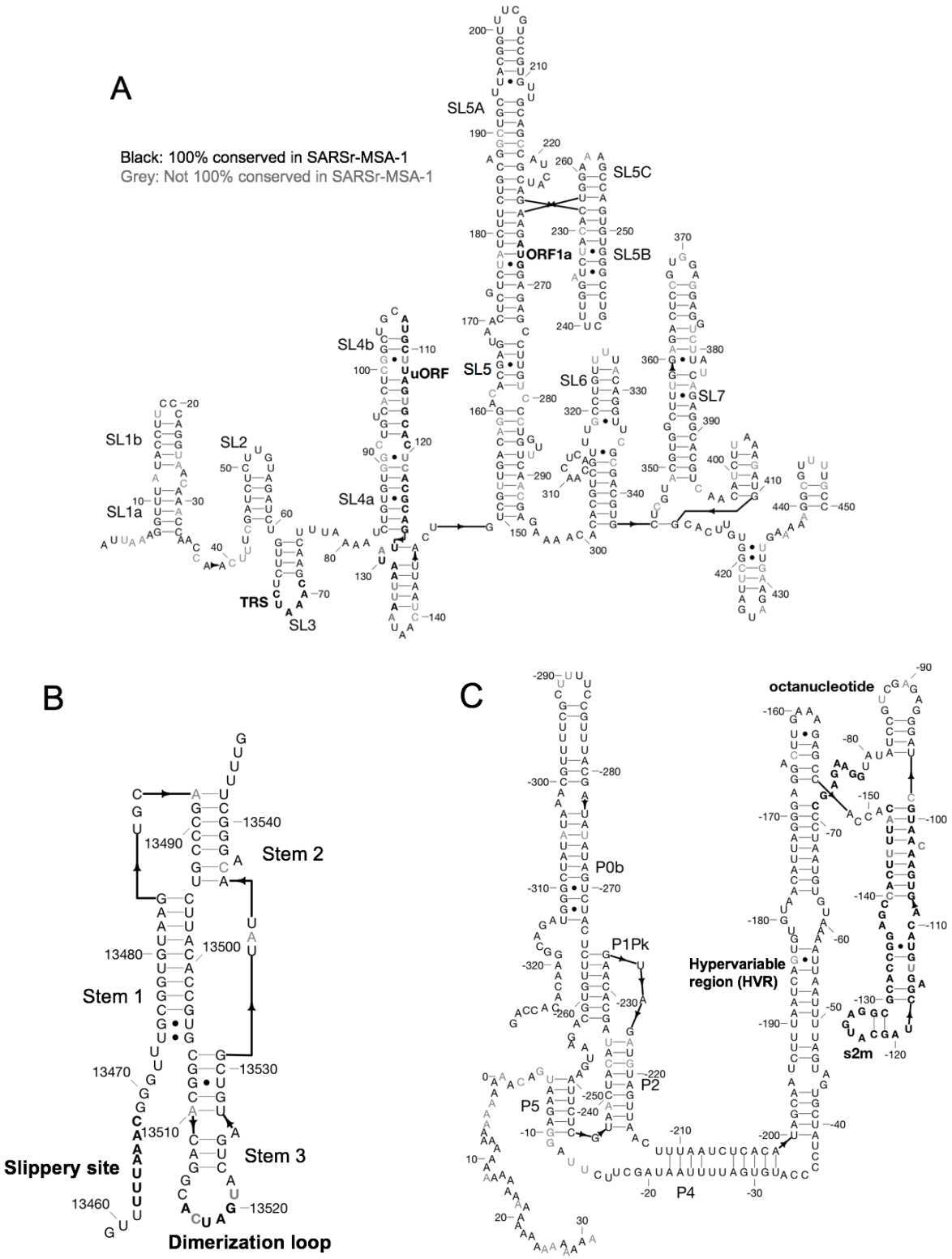
A
Black: 100% conserved in SARSr-MSA-1
Grey: Not 100% conserved in SARSr-MSA-1
SL5A
SL5C
SL5B
ORF1a
SL4b
uORF
SL5
SL6
SL7
SL1b
SL2
SL4a
SL1a
TRS
SL3
B
Stem 2
Stem 1
Stem 3
Slippery site
Dimerization loop
C
P0b
P1Pk
P2
P5
P4
octanucleotide
Hypervariable region (HVR)
s2m

Figure 4. Secondary structure diagrams for A) 5´ UTR, B) Frameshift element, C) 3´ UTR. Nucleotides are black if 100% conserved in the SARS, bat, and SARS-CoV-2 sequences in SARSr-MSA-1, and grey otherwise. Special labeled domains are in boldface. Structures are based primarily on manual identification of homology with literature coronavirus structure models. Note that numbering in (C) is relative to 3´ end of virus sequence. Figures prepared in RiboDraw (https://github.com/ribokit/RiboDraw).

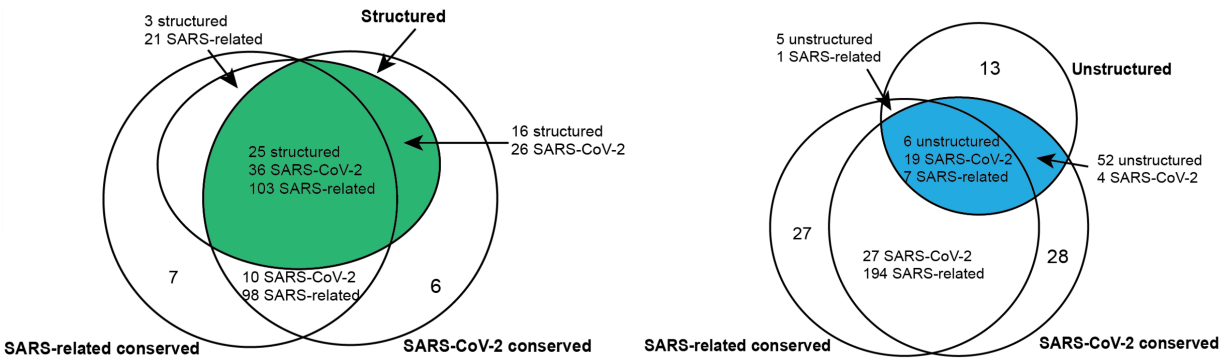

Figure 5. We depict the predicted number of structured, unstructured and conserved intervals for a choice of sequence conservation cutoffs. The SARS-related conserved intervals are all regions of at least 15 nucleotides with each position at least 90% conserved across an alignment of SARS, bat coronavirus, and SARS-CoV-2 sequences (SARSr-MSA-1). The SARS-CoV-2 intervals are regions of at least 15 nucleotides with each position at least 97% conserved across an alignment of currently available SARS-CoV-2 sequences (SARS-CoV-2-MSA-2). Structured intervals are *loci* predicted from RNAz with some *loci* containing multiple RNAz windows, and unstructured intervals are stretches of at least 15 nucleotides where all bases have base-pairing probability at most 0.4. All interval intersections are required to have at least 15 nucleotide overlaps, with the number of overlapping intervals listed for each interval type involved in the intersection. Top-scoring structured intervals conserved in SARS-CoV-2 sequences (green) are listed in Table 2. Top-scoring unstructured intervals conserved in SARS-CoV-2 sequences (blue) are listed in Table 3.

Name	Interval	Sequence	Conservation in SARS-CoV-2
SARS-related-conserved-1	14060-14075	UAGAUAAUCAAGAUCU	0.995896
SARS-related-conserved-2	15838-15857	UGGACUGAGACUGACCUUAC	0.995896
SARS-related-conserved-3	28554-28569	UUCGUGGUGGUGACGG	0.995868
SARS-related-conserved-4	28513-28546	AGAUGACCAAAUUGGCUACUACCGAAGAGCUACC	0.995868
SARS-related-conserved-5	16153-16169	GUUAUGCUUACUAAUGA	0.994528
SARS-related-conserved-6	27183-27212	GUACAGUAAGUGACAACAGAUGUUUCAUCU	0.99449
SARS-related-conserved-7	27165-27181	GACAAUAUUGCUUUGCU	0.99449
SARS-related-conserved-8	25511-25530	CACUCCCUUUCGGAUGGCUU	0.99449
SARS-related-conserved-9	25393-25409	AUGGAUUUGUUUAUGAG	0.99449
SARS-related-conserved-10	12905-12924	AGGUUUGUUACAGACACACC	0.99316
SARS-related-conserved-11	13346-13361	GUGGGUUUUACACUUA	0.99316
SARS-related-conserved-12	15496-15518	ACAACUGCUUAUGCUAAUAGUGU	0.99316
SARS-related-conserved-13	28799-28818	AGCAGAGGCGGCAGUCAAGC	0.993113
SARS-related-conserved-14	27457-27473	GAGUGUGUUAGAGGUAC	0.993113
SARS-related-conserved-15	25547-25562	UUCUUGCUGUUUUUCA	0.993113
SARS-related-conserved-16	17089-17105	GGUACUGGUAAGAGUCA	0.991792
SARS-related-conserved-17	17956-17975	UGCAUAAUGUCUGAUAGAGA	0.991792
SARS-related-conserved-18	18034-18050	UUACAAGCUGAAAAUGU	0.991792
SARS-related-conserved-19	704-723	GACGAGCUUGGCACUGAUCC	0.991792
SARS-related-conserved-20	25376-25392	UACACAUAAACGAACUU	0.991736
SARS-related-conserved-21	10406-10422	UACAAUGGUUCACCAUC	0.990437
SARS-related-conserved-22	16364-16388	UGUCUGUUAAUCCGUAUGUUUGCAA	0.990424
SARS-related-conserved-23	15622-15644	UAUGAGUGUCUCUAUAGAAAUAG	0.990424
SARS-related-conserved-24	15349-15367	GUUCUUGCUCGCAAACAUA	0.990424
SARS-related-conserved-25	15301-15323	AAAUGUGAUAGAGCCAUGCCUAA	0.990424
SARS-related-conserved-26	14077-14099	AAUGGUAACUGGUAUGAUUUCGG	0.990424
SARS-related-conserved-27	741-756	AAAACUGGAACACUAA	0.990424
SARS-related-conserved-28	25106-25128	GAAAUUGACCGCCUCAAUGAGGU	0.990358
SARS-related-conserved-29	26232-26267	UGAGUACGAACUUAUGUACUCAUUCGUUUCGGAAGA	0.990358
SARS-related-conserved-30	28270-28293	UAAAAUGUCUGAUAAUGGACCCCA	0.990358

Table 1: **SARS-related-conserved**. Conserved regions across SARSr-MSA-1 and SARS-CoV-2-MSA-2. All intervals are at least 90% conserved across the SARS and bat coronavirus sequences in SARSr-MSA-1, have length at least 15 nucleotides, and have every position at least 99% conserved in current GISAID SARS-CoV-2 sequences (SARS-CoV-2-MSA-2). Sequence intervals are relative to the reference genome NC_045512.2.

Name	Sequence interval	Sequence	Secondary structure	z-score	P value
SARS-CoV-2-conserved-structured-1	918-1037	ACUGGACUUUAUUGACACUAAGAGGGGUGUAUACUGCUGCC GUGAACAUGAGCAUGAAAUUGCUUGGUACACGGAACGUUCU GAAAAGAGCUAUGAAUUGCAGACACCUUUUGAAAUUAA	....................((((((((((...((((..(((((.((.((((((.....))))))))).)))))...(((((... .))))).....))))))))))))....	-4.38	0.999
SARS-CoV-2-conserved-structured-2	14206-14325	AUGUUGACACUGACUUAACAAAGCCUUACAUUAAGUGGGAU UUGUUAAAAUAUGACUUCACGGAAGAGAGGUUAAAACUCUU UGACCGUUAUUUUAAAUAUUGGGAUCAGACAUACCACC	(((((((..(..(.((((((((..((((.....))))))..))))))))).....((((.(((((.....)))))) ..)))).....).).)..))).)))).....	-1.99	0.999
SARS-CoV-2-conserved-structured-3	17125-17244	UCUACUACCCUUCUGCUCGCAUAGUGUAUACAGCUUGCUCU CAUGCCGCUGUUGAUGCACUAUGUGAGAAGGCAUUAAAAUA UUUGCCUAUAGAUAAAUGUAGUAGAAUUAUACCUGCAC	((((((((...((((((((((((((((((((..((.. ....)).)))))..)))))))))))))).(((((..... ..)))))..))))).....))))))).....	-6.85	0.999
SARS-CoV-2-conserved-structured-4	7011-7130	UUUAGGUGUUUUAAUGUCUAAUUUAGGCAUGCCUUCUUACU GUACUGGUUACAGAGAAGGCUAUUUGAACUCUACUAAUGUC ACUAUUGCAACCUACUGUACUGGUUCUAUACCUUGUAG	..(((((((..((((((((((...)))))))((((((..((((((.....))))))))))).....................))..)).)))))...(((.(((.....)))..))).	-1.36	0.999
SARS-CoV-2-conserved-structured-5	26135-26254	AAUCGACGGUUCAUCCGGAGUUGUUAAUCCAGUAAUGGAAC CAAUUUAUGAUGAACCGACGACGACUACUAGCGUGCCUUUG UAAGCACAAGCUGAUGAGUACGAACUUAUGUACUCAUUC	..(((.(((((((((..((((((....((((....)))))..))))))..)))))))))).))).....((((((((..... .))))..))))((((((((.....)))))))).	-5.44	0.999
SARS-CoV-2-conserved-structured-6	26175-26294	CCAAUUUAUGAUGAACCGACGACGACUACUAGCGUGCCUUU GUAAGCACAAGCUGAUGAGUACGAACUUAUGUACUCAUUCG UUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUA	....................(((..(((((.(((((((((((.... .(((....)))((((((((((.....)))))))))).(((((....)))))))))))))))..)))))).....))).	-4.44	0.999
SARS-CoV-2-conserved-structured-7	6251-6370	AUGCAACUAAUAAAGCCACGUAUAAACCAAAUACCUGGUGUA UACGUUGUCUUUGGAGCACAAAACCAGUUGAAACAUCAAAU UCGUUUGAGUGUACUGAAGUCAGAGGACGCGCAGGGAA	.(((.....((((((.(((((((.((((.....)))).)))))))).).)))))..((.....((((....((((((((... ..))))))))))))..(((....))))))).....	-2.52	0.999
SARS-CoV-2-conserved-structured-8	40-157	UUUCGAUCUCUUGUAGAUCUGUUCUCUAAACGAACUUUAAA AUCUGUGUGGCUGUCACUCGGCUGCAUGCUUAGUGCACUC ACGCAGUAUAAUUAAUAACUAAUUACUGUCGUUGACA	....((((.....)))))..........(((((.........(((((((..((.((((.(((.....))).)))))).))))))))..(((((.....)))))...))))).....	-2.91	0.998
SARS-CoV-2-conserved-structured-9	26491-26607	AGUUUUUCUGUUUGGAACUUUAAUUUUAGCCAUGGCAGAUU CCAACGGUACUAUUACCGUUGAAGAGCUUAAAAAGCUCCUU GAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCU	.((((.(((((((((..((.........)))))..)))))))... (((((((((....)))))))))..(((((.....)))))...))))...(((((((((....))))))))..........	-3.52	0.998
SARS-CoV-2-conserved-structured-10	25895-26014	CAUUACUUCAGGUGAUGGCACAACAAGUCCUAUUUCUGAAC AUGACUACCAGAUUGGUGGUUAUACUGAAAAAUGGGAAUCU GGAGUAAAAGACUGUGUUGUAUUACACAGUUACUUCAC	..(((((((((((..((.....))..(((((((((.(.(..((((((((((.....)))))))))..).).))))))))))))))))))..((((((((.....))))))).....	-4.4	0.998
SARS-CoV-2-conserved-structured-11	5211-5330	AUUAAAUCACACUAAAAAGUGGAAAUACCCACAAGUUAAUGG UUUAACUUCUAUUAAAUGGGCAGAUAACAACUGUUAUCUUG CCACUGCAUUGUUAACACUCCAACAAAUAGAGUUGAA	.....((.(((...(((.(.....))))...))).)).(((((((.((((.(.....((((((((((....)))))))..)))).....(((.....)))))))))))))))	-2.76	0.998
SARS-CoV-2-conserved-structured-12	26215-26334	UGUAAGCACAAGCUGAUGAGUACGAACUUAUGUACUCAUUC GUUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUACU UCUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUACAC	((((((((.((((.((((((((((.....)))))))))).)))).((((((((..((((((((((.....))))))))))))))))).)))).....(((((....)))))..))))).	-3.49	0.998
SARS-CoV-2-conserved-structured-13	6211-6330	GCUAAAUUGUUACAUAAACCUAUUGUUUGGCAUGUUAACAA UGCAACUAAUAAAGCCACGUAUAAACCAAAUACCUGGUGUA UACGUUGUCUUUGGAGCACAAAACCAGUUGAAACAUCA	(((..((((((((((...((.....)).)))).))))))) .....((((((.((((((.(((.....)))).))))))))).).)))))..)))......................	-1.03	0.998
SARS-CoV-2-conserved-structured-14	16445-16564	UUAUUGUAAAUCACAUAAACCACCCAUUAGUUUUCCAUUGU GUGCUAAUGGACAAGUUUUUGGUUUAUAUAAAAAUACAUGU GUUGGUAGCGAUAAUGUUACUGACUUUAAUGCAAUUGC	....(((((((((...((((...((((((((.....)))))))))....)))).)))))))).........(((..((. .((((.....))))))..))....)))....	-1.17	0.997
SARS-CoV-2-conserved-structured-15	7051-7170	UGUACUGGUUACAGAGAAGGCUAUUUGAACUCUACUAAUGU CACUAUUGCAACCUACUGUACUGGUUCUAUACCUUGUAGUG UUUGUCUUAGUGGUUUAGAUUCUUUAGACACCUAUCCU	(((.(((....))).(((((..((((((((...((....))(((((..((((((((.....(((.....)))..)))))).).))))..))))))))))))))))..)))....	0.99	0.996
SARS-CoV-2-conserved-structured-16	158-277	GGACACGAGUAACUCGUCUAUCUUCUGCAGGCUGCUUACG GUUUCGUCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGU UUCGUCCGGGUGUGACCGAAAGGUAAGAUGGAGAGCCUUG	.....(((...(((.(((((((.....(((((((.(((((. .....)))))..)))))))).........(((((((.((.....)))))))))(((....)))))))))))))).))))	-2.87	0.996
SARS-CoV-2-conserved-structured-17	23977-24096	UUACCAGAUCCAUCAAAACCAAGCAAGAGGUCAUUUAUUGA AGAUCUACUUUUCAACAAAGUGACACUUGCAGAUGCUGGCU UCAUCAAACAAUAUGGUGAUUGCCUUGGUGAUAUUGCU	((((((.....((((..(((((((.(((((((((.....)))))))..)))))))).)))))).....(((.(((((.....))))))..))))))))).....	-2.82	0.994
SARS-CoV-2-conserved-structured-18	25855-25974	ACUAUUGUAUACCUUACAAUAGUGUAACUUCUUCAAUUGUC AUUACUUCAGGUGAUGGCACAACAAGUCCUAUUUCUGAACA UGACUACCAGAUUGGUGGUUAUACUGAAAAAUGGGAAU	((((((((.....)))))))).....(((((((((.....)))))))))).....((((((((.(.(..((((((((.....)))))))))..).).).)))))))..	-4.25	0.994
SARS-CoV-2-conserved-structured-19	2943-3062	GGGCAUUGAUUUAGAUGAGUGGAGUAUGGCUACAUACUACU UAUUUGAUGAGUCUGGUGAGUUUAAAUUGGCUUCACAUAUG UAUUGUUCUUUCUACCCUCCAGAUGAGGAUGAAGAAGA	..((((((..(((((((((..((((((....)))))))))))))))((((..(...(.....)))..).)))))).)))) ....(((((.....(((.....))))..))))))..	-0.32	0.994
SARS-CoV-2-conserved-structured-20	14726-14845	UAAGGAAGGAAGUUCUGUUGAAUUAAAACACUUCUUCUUUG CUCAGGAUGGUAAUGCUGCUAUCAGCGAUUAUGACUACUAU CGUUAUAAUCUACCAACAAUGUGUGAUAUCAGACAACU	..((.(((((((...((((.....))))...)))))))).)).....(((((..(((....))))(((((((((.....))))))))))))....(((.((....)).)))...	-3.02	0.994
SARS-CoV-2-conserved-structured-21	26375-26490	CAAUAUUGUUAACGUGAGUCUUGUAAAACCUUCUUUUUACG UUUACUCUCGUGUUAAAAAUCUGAAUUCUUCUAGAGUUCCU GAUCUUCUGGUCUAAACGAACUAAAUAUUAUAUU	.(((((((....((.(((((.(((((((.....)))))))). ..)))).)))((.....(((((.....))))))..((((....))))..)))......)))))).....	-2.73	0.994
SARS-CoV-2-conserved-structured-22	21365-21484	AAUUCAGUUGUCUUCCUAUUCUUUAUUUGACAUAGAGUAAAU UUCCCCUUAAAUUAAGGGGUACUGCUGUUAUGUCUUUAAAA GAAGGUCAAAUCAAUGAUAUGAUUUUAUCUCUUCUUAG	..................((((((....(((((((((.....(((((((...)))))))...))))..))))))....)))))) (((.((((((.....)))))).)))..........	-2.27	0.994
SARS-CoV-2-conserved-structured-23	2158-2277	GAAGAGAAGUUUAAGGAAGGUGUAGAGUUUCUUAGAGACGG UUGGGAAAUUGUUAAAUUUAUCUCAACCUGUGCUUGUGAAA UUGUCGGUGGACAAAUUGUCACCUGUGCAAAGGAAAUU	.....((((.((((((.(....).)))))..))))))(((((((((((....))))..))))))))((..(..((((..(((((....)))))....))))..)..)).....	-1.62	0.993
SARS-CoV-2-conserved-structured-24	4174-4293	AAAAAGGCUGGUGGCACUACUGAAAUGCUAGCGAAAGCUUU GAGAAAAGUGCCAACAGACAAUUAUAUAACCACUUACCCGG GUCAGGGUUUAAAUGGUUACACUGUAGAGGAGGCAAAG	.....(.(((.(((((((.(.((.(.....((((....)))))).))...)))))))).))..((((((((((...((((... ...))))).....))))))...)))))..........	-1.33	0.993
SARS-CoV-2-conserved-structured-25	26528-26647	GAUUCCAACGGUACUAUUACCGUUGAAGAGCUUAAAAAGCU CCUUGAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCUUA CAUGGAUUUGUCUUCUACAAUUUGCCUAUGCCAACAGG	.....(((((((....)))))))..((((....)))) ((((.....((((((((.....))))))))..........(((((..((((.....))))....)).)))....))))	-3.29	0.993
SARS-CoV-2-conserved-structured-26	26255-26374	CGUUUCGGAAGAGACAGGUACGUUAAUAGUUAAUAGCGUAC UUCUUUUUCUUGCUUUCGUGGUAUUCUUGCUAGUUACACUA GCCAUCCUUACUGCGCUUCGAUUGUGUGCGUACUGCUG	.(((((.....)))))..((((((.((((((...((((((. .....(((.....))))).....((((.....))))) .....))))))..))))))..)))))))....	-2.31	0.992
SARS-CoV-2-conserved-structured-27	5651-5770	AUCUAGUACAACAGGAGUCACCUUUUGUUAUGAUGUCAGCA CCACCUGCUCAGUAUGAACUUAAGCAUGGUACAUUUACUUG UGCUAGUGAGUACACUGGUAAUUACCAGUGUGGUCACU	...((((((((((((....))))))))..(((((....((((..((((.(((....)))..)))))))))))).....))))))((((.((((((((((....))))))))).)))).	-2.77	0.992
SARS-CoV-2-conserved-structured-28	7131-7250	UGUUUGUCUUAGUGGUUUAGAUUCUUUAGACACCUAUCCUU CUUUAGAAACUAUACAAAUUACCAUUUCAUCUUUUAAAUGG GAUUUAACUGCUUUUGGCUUAGUUGCAGAGUGGUUUUU	.((((((..(((((.((((((...)))))))..(((..... ..)))..)))).)))))..(((((.....)))))). ...(((..(((...(.....)).)))..)))...	-0.12	0.992
SARS-CoV-2-conserved-structured-29	10970-11089	AAAGUGCAGUGAAAAGAACAAUCAAGGGUACACACCACUGG UUGUUACUCACAAUUUUGACUUCACUUUUAGUUUUAGUCCA GAGUACUCAAUGGUCUUUGUUCUUUUUUUUGUAUGAAA	...((((((.((((((((((.(((..(((....)))..)))((((.(((((...(((..(((.....)))..)))...))))))..))))).....))))))))))).))))))....	-1.44	0.992

SARS-CoV-2-conserved-structured-30	28838-28957	GUAGUCGCAACAGUUCAAGAAAUUCAACUCCAGGCAGCAGU AGGGGAACUUCUCCUGCUAGAAUGGCUGGCAAUGGCGGUG AUGCUGCUCUUGCUUUGCUGCUGCUUGACAGAUUGAACC	.(((((.....((((.(....)).)))).(((((((((((((((.....)))))))).....(((.(((((.((((.... .)))))).)))))..)))))))))))...))))))....	-2.64	0.992
SARS-CoV-2-conserved-structured-31	26095-26214	AAAUUGUUGAUGAGCCUGAAGAACAUGUCCAAAUUCACACA AUCGACGGUUCAUCCGGAGUUGUUAAUCCAGUAAUGGAACC AAUUUAUGAUGAACCGACGACGACGACUACUAGCGUGCCUU	.....(((.((..(.(.(.....)).).)).)).....(((.((((((((((..(((((((....((((....)))).))))))).)))))))))).))))))).....)	-1.91	0.992
SARS-CoV-2-conserved-structured-32	2903-3022	UAAAAACUUUGCAACCAGUAUCUGAAUUACUUACACCACUG GGCAUUGAUUUAGAUGAGUGGAGUAUGGCUACAUACUACUU AUUUGAUGAGUCUGGUGAGUUUAAAUUGGCUUCACAUA	.....((((((.....)))((((.((((..(((.((...((((((((((..(((((((....)))))))))))))))).))))))..)))..)))..))).)))..))).....)	-0.97	0.991
SARS-CoV-2-conserved-structured-33	29550-29669	AAGGCAGAUGGGCUAUAUAAACGUUUUCGCUUUUCCGUUUA CGAUAUAUAGUCUACUCUUGUGCAGAAUGAAUUCUCGUAAC UACAUAGCACAAGUAGAUGUAGUUAACUUUAAUCUCAC	.(((.((.((((((((((..(((...((.....)).)...))).))))))))))))))((.(((.((((.....((((((((.((....))..)))))))))..)))).))))))	-1.83	0.990
SARS-CoV-2-conserved-structured-34	80-197	AAUCUGUGUGGCUGUCACUCGGCUGCAUGCUUAGUGCACU CACGCAGUAUAAUUAAUAACUAAUUACUGUCGUUGACAGGA CACGAGUAACUCGUCUAUCUUCUGCAGGCUGCUUACG	...(((((((..((.(((.(((.....))).))))))..)))))))..(((((.....)))))..((.(((..(((((.(((((...)))))).....)))))..)))..)).).....	-0.8	0.990
SARS-CoV-2-conserved-structured-35	13886-14005	CAAUUGUUGUGAUGAUGAUUAUUUCAAUAAAAAGGACUGGU AUGAUUUUGUAGAAAACCCAGAUAUAUUACGCGUAUACGCC AACUUAGGUGAACGUGUACGCCAAGCUUUGUUAAAAAC	...((((((.((((((...)))))).))))))((((..((((. .....))))..........((((((((((...((....))...)))))))))....))))..........	-0.37	0.989
SARS-CoV-2-conserved-structured-36	120-237	CACGCAGUAUAAUUAAUAACUAAUUACUGUCGUUGACAGGA CACGAGUAACUCGUCUAUCUUCUGCAGGCUGCUUACGGUU UCGUCCGUGUUGCAGCCGAUCAUCAGCACAUCUAGGU	...((((((..((((.....)))))))))).((((((((((.((((((...)))))).....)))))..((((((.((((..... ..)))))..)))))).....))))).....	-2.33	0.989
SARS-CoV-2-conserved-structured-37	23697-23816	UGCCAUACCCACAAAUUUUACUAUUAGUGUUACCACAGAAA UUCUACCAGUGUCUAUGACCAAGACAUCAGUAGAUUGUACA AUGUACAUUUGUGGUGAUUCAACUGAAUGCAGCAAUCU	.....((((((((((((((((...(((((..((((((.....))))))..)))))..(((((...))))))))))))))))...))))).....	-2.64	0.989
SARS-CoV-2-conserved-structured-38	25655-25774	AACAGUUUACUCACACCUUUUGCUCGUUGCUGCUGGCCUU GAAGCCCCUUUUCUCUAUCUUUAUGCUUUAGUCUACUUCUU GCAGAGUAUAAACUUUGUAAGAAUAAUAAUGAGGCUUUG	..((((.....((.....)).)))((((((((((.....)))))).....(((((((((((....)))))))))))).....))))))....	-1.15	0.989
SARS-CoV-2-conserved-structured-39	15606-15724	UUACAACACAGACUUUAUGAGUGUCUCUAUAGAAAUAGAGA UGUUGACACAGACUUUGUGAAUGAGUUUUACGCAUAUUUGC GUAAACAUUUCUCAAUGAUGAUACUCUCUGACGAUGC	.....(((.....((((((..(((.((....(((((((..(((((...)))))).....((((((....))))))))))))))))).)))..)))))))))).....)	-1.81	0.988
SARS-CoV-2-conserved-structured-40	14526-14645	AGCUCUAGACUUAGUUUUAAGGAAUUACUUGUGUAUGCUGC UGACCCUGCUAUGCACGCUGCUUCUGGUAAUCUAUUACUAG AUAAACGCACUACGUGCUUUUCAGUAGCUGCACUUACU	.....((((((....))))))...(((((.((((((((.....(((((.(((.((((((((....)))))))). ....)))...)))))...))))))))))))).....	-1.64	0.986
SARS-CoV-2-conserved-structured-41	3808-3897	CUCUAUGACAAACUUGUUUCAAGCUUUUUGGAAAUGAAGAG UGAAAAGCAAGUUGAACAAAAGAUCGCUGAGAUUCCUAAAG AGGAAGUU	(((..(((..((((((((((...((((((.....)))))))...)))))))).....)))..))).((((((....)))))...	-1.32	0.983
SARS-CoV-2-conserved-structured-42	25935-26054	CAUGACUACCAGAUUGGUGGUUAUACUGAAAAAUGGGAAUC UGGAGUAAAAGACUGUGUUGUAUUACACAGUUACUUCACUU CAGACUAUUACCAGCUGUACUCAACUCAAUUGAGUACA	.(((((((((.....)))))))))..(((...(((((...((((((((((..((((((((.....)))))))))..)).))))))))))))..)))..((((((((.....)))))))).	-4.97	0.983
SARS-CoV-2-conserved-structured-43	15645-15764	GAUGUUGACACAGACUUUGUGAAUGAGUUUUACGCAUAUUU GCGUAAACAUUUCUCAAUGAUGAUACUCUCUGACGAUGCUG UUGUGUGUUUCAAUAGCACUUAUGCAUCUCAAGGUCUA	(((.((((..((((..(..(.(.((((.(((((((....))))))))).....)))).).).)..).....)))).((((... ((.((((.....)))))..)).))))))))).)))..	-2.07	0.983
SARS-CoV-2-conserved-structured-44	2518-2625	UUAACAGAGGAAGUUGUCUUGAAAACUGGUGAUUUACAACC AUUAGAACAACCUACUAGUGAAGCUGUUGAAGCUCCAUUGG UUGGUACACCAGUUUGUAUUAACGGG	.....((((.....(((((((((.(((.....))))...(((.(((((((.(((.....)))).)))))))).))).))))))))....))))....	-1.93	0.982
SARS-CoV-2-conserved-structured-45	26568-26687	UCCUUGAACAAUGGAACCUAGUAAUAGGUUUCCUAUUCCUU ACAUGGAUUUGUCUUCUACAAUUUGCCUAUGCCAACAGGAA UAGGUUUUUGUAUAUAAUUAAGUUAAUUUUCCUCUGGC	..((((((((((.((((((((....)))))))).....((((.. ....))).))))....(((((...((((((.((....)).)))))))..)))))).....))))).....	-2.13	0.982
SARS-CoV-2-conserved-structured-46	1238-1357	AUUGUGGUGAAACUUCAUGGCAGACGGGCGAUUUUGUUAAA GCCACUUGCGAAUUUUGUGGCACUGAGAAUUUGACUAAAGA AGGUGCCACUACUUGUGGUUACUUACCCCAAAAUGCUG	.((((.((((....)))).)))))...(((.(((((.....((((.(.....)))))).....(((((.....)))).).)))))).....)))).....))))))))).	-0.75	0.981
SARS-CoV-2-conserved-structured-47	6451-6570	GAGUGUAAUGUGAAAACUACCGAAGUUGUAGGAGACAUUAU ACUUAAACCAGCAAAUAAUAGUUUAAAAAUUACAGAAGAGGU UGGCCACACAGAUCUAAUGGCUGCUUAUGUAGACAAU	((((((((((.....((((.....)))).....))))))))))).....(((((....((((.(((((.....)))))))))).)))).....	-0.83	0.981
SARS-CoV-2-conserved-structured-48	29078-29197	AUGUAACACAAGCUUUCGGCAGACGUGGUCCAGAACAAACC CAAGGAAAUUUUGGGGACCAGGAACUAAUCAGACAAGGAAC UGAUUACAAACAUUGGCCGCAAAUUGCACAAUUUGCCC	.....(((...(.((((((.(.....)(((((.....)))))))))).)...((((((.....))))))))).....)))((((((....))))))..	-1.24	0.980
SARS-CoV-2-conserved-structured-49	23497-23616	CGUGCAGGCUGUUUAAUAGGGGCUGAACAUGUCAACAACUC AUAUGAGUGGACAUACCCAUUGGUGCAGGUAUAUGCGCUA GUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGU	(((((..((((.....(((.((((..(((((.... ...)))).)))))))....((((((((.....)))))))))..))))..)))(((.....)))...)))))..)))))).	-2.42	0.979
SARS-CoV-2-conserved-structured-50	13686-13805	GAAGAAACAAUUUAUAAUUUACUUAAGGAUUGUCCAGCUGU UGCUAAACAUGACUUCUUUAAGUUUAGAAUAGACGGUGACA UGGUACCACAUAUAUCACGUCAACGUCUUACUAAAUAC	.....(((((((((.(((..(((....)))..)))).....)))))))))(((...((((.((((.((((((.....)))))).)))).)))))..))))....	-3.1	0.978
SARS-CoV-2-conserved-structured-51	12610-12729	AUGGACAAUUCACCUAAUUUAGCAUGGCCUCUUAUUGUAAC AGCUUUAAGGGCCAAUUCUGCUGUCAAAUUACAGAAUAAUG AGCUUAGUCCUGUUGCACUACGACAGAUGUCUUGUGCU	..((((((((((.....(((((.((((..((....))..)))))))))((((((.((.....)))))))))..)))) .)).)))....(((...(((....))).)).))).	-1.94	0.978
SARS-CoV-2-conserved-structured-52	19645-19764	UUGUAAAUAAGGGACACUUUGAUGGACAACAGGGUGAAGUA CCAGUUUCUAUCAUUAAUAACACUGUUUACACAAAAGUUGA UGGUGUUGAUGUAGAAUUGUUUGAAAAUAAAACAACAU	...(((((((...(((...(((((((.((.((.....)).)))))))))....((((((((((.((.....)).))))))))))).))).....))))))).....	-0.89	0.977
SARS-CoV-2-conserved-structured-53	11050-11169	CAGAGUACUCAAUGGUCUUUGUUCUUUUUUUUGUAUGAAAA UGCCUUUUUACCUUUUGCUAUGGGUAUUAUUGCUAUGUCU GCUUUUGCAAUGAUGUUUGUCAAACAUAAGCAUGCAUUU	(((((.(((.....)))))))).....(((((((..... ((..(..((((.....)))))..)..))(((((.(((. ...))).(((.....))).)))))..)))))))...	-0.33	0.977
SARS-CoV-2-conserved-structured-54	17445-17564	CAAUUACCUGCACCACGCACAUUGCUAACUAAGGGCACACU AGAACCAGAAUAUUUCAAUUCAGUGUGUAGACUUAUGAAAA CUAUAGGUCCAGACAUGUUCCUCGGAACUUGUCGGCGU	.....((.((((((((..((((.....)))))..... .....(((.....)))).)))))).(((((((.....))))))))..(((.((((....))).)))))..)))))).)	-2.14	0.974
SARS-CoV-2-conserved-structured-55	24256-24375	GCUAUGCAAAUGGCUUAUAGGUUUAAUGGUAUUGGAGUUAC ACAGAAUGUUCUCUAUGAGAACCAAAAAUUGAUUGCCAACC AAUUUAAUAGUGCUAUUGGCAAAAUUCAAGACUCACUU	(((((....)))))...................(((((..... ....((((((...))))))).....((((((((((...(((....)))....)))))))...))))))))....	-0.32	0.974
SARS-CoV-2-conserved-structured-56	17525-17644	AACUAUAGGUCCAGACAUGUUCCUCGGAACUUGUCGGCGUU GUCCUGCUGAAAUUGUUGACACUGUGAGUGCUUUGGUUUA UGAUAAUAAGCUUAAAGCACAUAAAGACAAAUCAGCUCA	.....(((.((.((((.((((....)))).))))))..... .)))(((((..(((((.....((.(((((((((((((. ...)))))).))))))))..)))..)))))..))))...	-2.22	0.973
SARS-CoV-2-conserved-structured-57	27248-27367	AAUUAUUAUGAGGACUUUUAAAGUUUCCAUUUGGAAUCUUG AUUACAUCAUAAACCUCAUAAUUAAAAAUUUAUCUAAGUCAC UAACUGAGAAUAAAUAUUCUCAAUUAGAUGAAGAGCA	....((((((((((.....(((.(((((....)))).)))(((...)))).....)))))))))..........((...(((((((.(((((((.....)))))).))))).))).))..))..	-3.72	0.972
SARS-CoV-2-conserved-structured-58	4574-4693	CACUUAACGAUCUAAAUGAAACUCUUGUUACAAUGCCACUU GGCUAUGUAACACAUGGCUUAAAUUUGGAAGAAGCUGCUCG GUAUAUGAGAUCUCUCAAAGUGCCAGCUACAGUUUCUG	.....((((((((..((..(.(((((((..((.....)))..)))))))...)..)).)))))))(((((((.(((((.((((.....)))).)))))))...)))))).	-3.73	0.971
SARS-CoV-2-conserved-structured-59	21325-21444	UCAUGCAUGCAAAUUACAUAUUUUGGAGGAAUACAAAUCCA AUUCAGUUGUCUUCCUAUUCUUUAUUUGACAUGAGUAAAUU UCCCCUUAAAUUAAGGGGUACUGCUGUUAUGUCUUUAA	.((((((((((((.....((...(((((((..(..... ....))..)))))).))).....)))))).))))(((.....(((((((...)))))))...))))..))))).....	-1.24	0.970

SARS-CoV-2-conserved-structured-60	7771-7890	CUUUACUUUGAUAAAGCUGGUCAAAAGACUUAUGAAAGACA UUCUCUCUCUCAUUUUGUUAACUUAGACAACCUGAGAGCUA AUAACACUAAAGGUUCAUUGCCUAUUAAUGUUAUAGUU	.....(((.(((..((.....))..)))).))))..... ....((((((..((((.....)))))..))))))...((((((.(((((((.....)))).))).))))))....	-1.93	0.967
SARS-CoV-2-conserved-structured-61	19765-19884	UACCUGUUAAUGUAGCAUUUGAGCUUUGGGCUAAGCGCAAC AUUAAACCAGUACCAGAGGUGAAAAUACUCAAUAAUUUGGG UGUGGACAUUGCUGCUAAUACUGUGAUCUGGGACUACA	..(((.(((((((.((.(((.(((.....)))))).)).)).)))))))).(((((.(((..(((...(((((((....)))))))...)))..)))....)))))).).....))).....	-0.84	0.965
SARS-CoV-2-conserved-structured-62	6971-7090	UAAGUGUUUGCCUAGGUUCUUUAAUCUACUCAACCGCUGCU UUAGGUGUUUUAAUGUCUAAUUUAGGCAUGCCUUCUUACUG UACUGGUUACAGAGAAGGCUAUUUGAACUCUACUAAUG	..((((.(((..(((((....)))))))..)))).)))).... ((((.(((..((((((((...)))))))(((((((..((((((.....)))))))))))).....))))))).....	-1.9	0.963
SARS-CoV-2-conserved-structured-63	28998-29117	AGGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCUA AGAAGCCUCGGCAAAAACGUACUGCCACUAAAGCAUACAAU GUAACACAAGCUUUCGGCAGACGUGGUCCAGAACAAAC	.((((.....((((((((((...))))))))))))...(((.(((((...(((((..... .....))))).)))))))))))).....	-4.19	0.963
SARS-CoV-2-conserved-structured-64	5011-5130	GUUGUGGACAUGUCAAUGACAUAUGGACAACAGUUUGGUCC AACUUAUUUGGAUGGAGCUGAUGUUACUAAAAUAAAACCUC AUAAUUCACAUGAAGGUAAAACAUUUUAUGUUUUACCU	..((((((.(((((...)))))...((((.((((((((((.....)))))).))))))).)))).................. ..)))))...((((((((((...)))))))))))	-4.19	0.963
SARS-CoV-2-conserved-structured-65	4094-4213	CUUUCUUAAAGAAAGAUGCUCCAUAUAUAGUGGGUGAUGUU GUUCAAGAGGGUGUUUUAACUGCUGUGGUUAUACCUACUAA AAAGGCUGGUGGCACUACUGAAAUGCUAGCGAAAGCUU	..(((((.((.((...(((.(((.....)))))))...)).)).)))))((((((((.....((((((..(((..((.... .))..))))..)))))))))))))))((....)).	0.4	0.958
SARS-CoV-2-conserved-structured-66	1598-1717	GUCUUAAUGACAACCUUCUUGAAAUACUCCAAAAAGAGAAAG UCAACAUCAAUAUUGUUGGUGACUUUAAACUUAAUGAAGAG AUCGCCAUUAUUUUGGCAUCUUUUUCUGCUUCCACAA	(((.....)))...(((((((((....(((.....)))((((((((.((.((....)).)).)))))))....)))).)))))((((.(((.....))))))).....	-2	0.955
SARS-CoV-2-conserved-structured-67	11090-11209	AUGCCUUUUUACCUUUUGCUAUGGGUAUUAUUGCUAUGUCU GCUUUUGCAAUGAUGUUUGUCAAACAUAAGCAUGCAUUUCU CUGUUUGUUUUUGUUACCUUCUCUUGCCACUGUAGCUU	.....(((((((.((((((((((...(....) ....)))))))).....((((.(((((.(((.....))).)))))..))))..............))))))))..	-1.63	0.955
SARS-CoV-2-conserved-structured-68	17965-18084	CUGAUAGAGACCUUUAUGACAAGUUGCAAUUUACAAGCUU GAAAUUCCACGUAGGAAUGUGGCAACUUUACAAGCUGAAAA UGUAACAGGACUCUUUAAAGAUUGUAGUAAGGUAAUCA	..(((....((((....((((((((((.(((...((.((((((((((((((....)))))).)).)))..)))))).)))..))))))).....(((....)))))))...)))))..))).	-0.96	0.952
SARS-CoV-2-conserved-structured-69	25815-25934	GAUGCCAACUAUUUUCUUUGCUGGCAUACUAAUUGUUACGA CUAUUGUAUACCUUACAAUAGUGUAACUUCUUCAAUUGUCA UUACUUCAGGUGAUGGCACAACAAGUCCUAUUUCUGAA	.((((((.(..........).)))))).....(((((.((((((((.....)))))))))))))).....((((((((((.....))))))))....	-4.17	0.952
SARS-CoV-2-conserved-structured-70	14446-14565	AUGGUGUUCCAUUUGUAGUUUCAACUGGAUACCACUUCAGA GAGCUAGGUGUUGUACAUAAUCAGGAUGUAAACUUACAUAG CUCUAGACUUAGUUUUAAGGAAUUACUUGUGUAUGCUG	..(((((..(((..((((((((..(((((.....)))))(((((((.(((...(((((.....)))))....))).)))))))).....))))))))..))).)))))).	-1.41	0.948
SARS-CoV-2-conserved-structured-71	2118-2237	CACUGUUUAUGAAAAACUCAAACCCGUCCUUGAUUGGCUUG AAGAGAAGUUUAAGGAAGGUGUAGAGUUCUUAGAGACGGU UGGGAAAUUGUUAAAUUUAUCUCAACCUGUGCUUGUGA	(((.((((.(((.((((((..(((..(((((((...((((. .....)))))))))))).))...)))))))..))).)))))((((((((((((.....)))).)))))))).....))).	-2.75	0.948
SARS-CoV-2-conserved-structured-72	17205-17324	UAUUUGCCUAUAGAUAAAUGUAGUAGAAUUAUACCUGCACG UGCUCGUGUAGAGUGUUUUGAUAAAUUCAAAGUGAAUUCAA CAUUAGAACAGUAUGUCUUUUGUACUGUAAAUGCAUUG	((((((....))))))(((((((.(((((((((((((((....)))))))..))))((((.....))))...))))))).)).....(((((((.....)))))))...)))))).	-2.61	0.947
SARS-CoV-2-conserved-structured-73	7171-7290	UCUUUAGAAACUAUACAAAUUACCAUUUCAUCUUUUAAAUGG GAUUUAACUGCUUUUGGCUUAGUUGCAGAGUGGUUUUUGG CAUAUAUUCUUUUCACUAGGUUUUUCUAUGUACUUGGA	...((((((((((..(((((.((((((.....))))))))))..((((..(((...)))..)))).)))))))))) .....(((((.....))))).	-0.53	0.945
SARS-CoV-2-conserved-structured-74	14766-14885	GCUCAGGAUGGUAAUGCUGCUAUCAGCGAUUAUGACUACUA UCGUUAUAAUCUACCAACAAUGUGUGAUAUCAGACAACUAC UAUUUGUAGUUGAAGUUGUUGAUAAGUACUUUGAUUGU	(((...(((((((....)))))))))))((((((((..... ..)))))))))...(((((((((..((....)).))))))) .....))))))...))))))).....	-2.13	0.943
SARS-CoV-2-conserved-structured-75	11730-11849	UAUGAAUUCACAGGGACUACUCCCACCCAAGAAUAGCAUAG AUGCCUUCAAACUCAACAUUAAAUUGUUGGGUGUUGGUGGC AAACCUUGUAUCAAAGUAGCCACUGUACAGUCUAAAAU	...((....(((((((.(((((.....((((.....((((.(((((((((((.....)))))))).)))).))))). ..))))).....)))))))..)))....)).....	-1.59	0.942
SARS-CoV-2-conserved-structured-76	29238-29357	GGAAGUCACACCUUCGGGAACGUGGUUGACCUACACAGGU GCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAAGAUCAA GUCAUUUUGCUGAAUAAGCAUAUUGACGCAUACAAAAC	....(((((.(((..((....)).)))))))).....((((. .....((((((.....))))))).....((((. ..(((.....)))).....))))))).....	-1.58	0.941
SARS-CoV-2-conserved-structured-77	1558-1677	GGUUGUAACCAUACAGGUGUUGUUGGAGAAGGUUCCGAAG GUCUUAAUGACAACCUUCUUGAAAUACUCCAAAAAGAGAAAG UCAACAUCAAUAUUGUUGGUGACUUUAAACUUAAUGAA	(((....)))).....((.((..((.(((((((((.....((((.....)))))))))))))..)).)..)).)...(((..(((((((.((.((....)).)).)).))))))...)))......	-1.18	0.941
SARS-CoV-2-conserved-structured-78	4214-4333	UGAGAAAAGUGCCAACAGACAAUUAUAUAACCACUUACCCG GGUCAGGGUUUAAAUGGUUACACUGUAGAGGAGGCAAAGAC AGUGCUUAAAAAGUGUAAAAGUGCCUUUUACAUUCUAC	..((((.((((.....(((((((...((((... ...))))).....)))))))))))((((..(((((...(((. (.....).)))).....)))))))))..)))..	-0.39	0.940
SARS-CoV-2-conserved-structured-79	9410-9529	CUGGUGGUAUUGUAGCUAUCGUAGUAACAUGCCUUGCCUA CUAUUUUAUGAGGUUUAGAAGAGCUUUUGGUGAAUACAGUC AUGUAGUUGCCUUUAAUACUUUACUAUUCCUUAUGUCAU	..(((((((((.((.((((...)))).)))))))..))).... .....(((((((..(((.((((..((..(.((((....))))).)..))...))..)))).))).)))))))....	-1.74	0.940
SARS-CoV-2-conserved-structured-80	13806-13925	ACAAUGGCAGACCUCGUCUAUGCUUUAAGGCAUUUUGAUGA AGGUAAUUGUGACACAUUAAAAGAAAUACUUGUCACAUACAA UUGUUGUGAUGAUGAUUAUUUCAAUAAAAAGGACUGG	(((((.....((((((((.((((....)))))))...)))).)))).)))))..........((((((.(..(((((((((.. ..).)))).)))..).)))))).....	-1.73	0.940
SARS-CoV-2-conserved-structured-81	23537-23656	CAUAUGAGUGUGACAUACCCAUUGGUGCAGGUAUAUGCGCU AGUUAUCAGACUCAGACUAAUUCUCCUCGGCGGGCACGUAG UGUAGCUAGUCAAUCCAUCAUUGCCUACACUAUGUCAC	....(((((.(((.....((((((((.....)))))))))..)))..)))))..........((.....))(.((((((((((.(((.(.....)))).))))))))))..	-0.51	0.940
SARS-CoV-2-conserved-structured-82	9210-9329	GUACUGUAGGCACGGCACUUGUGAAAGAUCAGAAGCUGGU GUUUGUGUAUCUACUAGUGGUAGAUGGGUACUUAACAAUGA UUAUUACAGAUCUUUACCAGGAGUUUUCUGUGGUGUAGA	(((((.(((((((.(..(((.(((....))).)))).).)))))))..((((((.....)))))))))))..........(((((((((.(((....))))...))))))))....	-0.31	0.939
SARS-CoV-2-conserved-structured-83	6091-6210	CAGUUAACUGGUUAUAAGAAACCUGCUUCAAGAGAGCUUAA AGUUACAUUUUUCCCUGACUUAAAUGGUGAUGUGGUGGCUA UUGAUUAUAAACACUACACACCCUCUUUUAAGAAAGGA	..((.((((((((.....)))).((((.....))))....))))))).....(((..((((((.(((.(((((((..(((.....)))..)))))))))))....))))))..))).	-2.57	0.939
SARS-CoV-2-conserved-structured-84	29118-29237	CCAAGGAAAUUUUGGGGACCAGGAACUAAUCAGACAAGGAA CUGAUUACAAACAUUGGCCGCAAAUUGCACAAUUUGCCCCC AGCGCUUCAGCGUUCUUCGGAAUGUCGCGCAUUGGCAU	((((.....(((((.(((.....((((((((.....)))))))).....)))))..((((((.....))))))))))))(((((....(((((.(....)))))).)))).))))...	-3.14	0.938
SARS-CoV-2-conserved-structured-85	28958-29077	AGCUUGAGAGCAAAAUGUCUGGUAAAGGCCAACAACAACAA GGCCAAACUGUCACUAAGAAAUCUGCUGCUGAGGCUUCUAA GAAGCCUCGGCAAAAACGUACUGCCACUAAAGCAUACA	.((((.((.((..(((((..(((((.....)))))...((.....)))))).....))))..((((((((((...))))))))))).....))))..))).))).)))).....	-2.78	0.935
SARS-CoV-2-conserved-structured-86	29278-29397	GCCAUCAAAUUGGAUGACAAAGAUCCAAAUUUCAAAGAUCAA GUCAUUUUGCUGAAUAAGCAUAUUGACGCAUACAAAACAUU CCCACCAACAGAGCCUAAAAAGGACAAAAAGAAGAAG	.....(((((.....))))))..(((.....((((...(((.....)))).....)))).................. .....(((....))).).....))))...	-1.58	0.935
SARS-CoV-2-conserved-structured-87	14046-14165	GGUGUACUGACAUUAGAUAAUCAAGAUCUCAAUGGUAACUG GUAUGAUUUCGGUGAUUUCAUACAAACCACGCCAGGUAGUG GAGUUCCUGUUGUAGAUUCUUAUUAUUCAUUGUUAAUG	.....(((((.(..(((((.((((((((((((.(((.(. ((((((..((...)).)))))))...(((.....)))))))).))))))).))).))))))))..).)))))....	-2.05	0.935
SARS-CoV-2-conserved-structured-88	9250-9369	GUUUGUGUAUCUACUAGUGGUAGAUGGGUACUUAACAAUGA UUAUUACAGAUCUUUACCAGGAGUUUUCUGUGGUGUAGAUG CUGUAAAUUUACUUACUAAUAUGUUUACACCACUAAUU	(((...((((((((((....)))).)))))))..))))..... ....((((.(((....)))).....))))(((((((((... (((.....)))).....)))))))).....	-0.53	0.934
SARS-CoV-2-conserved-structured-89	26415-26527	GUUUACUCUCGUGUUAAAAAUCUGAAUUCUUCUAGAGUUCC UGAUCUUCUGGUCUAAACGAACUAAAUAUUAUAUUAGUUUU UCUGUUUGGAACUUUAAUUUUAGCCAUGGCA	.....(((((((((((....((.....))((((((((. ((((....)))).((((((((((.....)))))))... .)))))))))))).))))))).)))))..	-1.94	0.934

SARS-CoV-2-conserved-structured-90	14846-14965	ACUAUUUGUAGUUGAAGUUGUUGAUAAGUACUUUGAUUGUU ACGAUGGUGGCUGUAUUAAUGCUAACCAAGUCAUCGUCAAC AACCUAGACAAAUCAGCUGGUUUUCCAUUUAAUAAAUG	.....((((((...((((((..((.((..(.(((((.. ((((((((((...(((....)))...)))))))))))))))..))))..))))))(((....))))))))..... .

Name	Sequence interval	Average unpaired probability	Minimum unpaired probability	Sequence
SARS-CoV-2-conserved-unstructured-1	29074-29087	0.891105	0.763759	AUACAAUGUAACAC
SARS-CoV-2-conserved-unstructured-2	8078-8094	0.82497	0.752542	CCAAUGGAAAAACUCAA
SARS-CoV-2-conserved-unstructured-3	1359-1374	0.836512	0.716868	UUGUUAAAAUUUAUUG
SARS-CoV-2-conserved-unstructured-4	21626-21643	0.857235	0.713387	ACUCAAUUACCCCCUGCA
SARS-CoV-2-conserved-unstructured-5	1420-1436	0.796631	0.697304	CGAAUACCAUAAUGAAU
SARS-CoV-2-conserved-unstructured-6	18471-18484	0.779821	0.695284	UCAAUUUAAACACC
SARS-CoV-2-conserved-unstructured-7	11910-11923	0.766973	0.683262	AAUCAUCAUCUAAA
SARS-CoV-2-conserved-unstructured-8	23960-23981	0.787037	0.677563	UUUAAUUUUUCACAAAUAUUAC
SARS-CoV-2-conserved-unstructured-9	13990-14003	0.796334	0.661868	CAAGCUUUGUUAAA
SARS-CoV-2-conserved-unstructured-10	10009-10035	0.760204	0.656672	UCUUUACCAACCACCACAAACCUCUAU
SARS-CoV-2-conserved-unstructured-11	23700-23718	0.822733	0.654787	CCAUACCCACAAAUUUUAC
SARS-CoV-2-conserved-unstructured-12	18918-18934	0.832132	0.654059	UAUUGAAUAUCCUAUAA
SARS-CoV-2-conserved-unstructured-13	27385-27402	0.809693	0.653733	UAAACGAACAUGAAAAUU
SARS-CoV-2-conserved-unstructured-14	5773-5789	0.808284	0.65353	UAAACAUAUAACUUCUA
SARS-CoV-2-conserved-unstructured-15	23910-23932	0.837777	0.652669	CACAAGUCAAACAAAUUUACAAA
SARS-CoV-2-conserved-unstructured-16	17762-17785	0.767426	0.650277	CUGUCUUUAUUUCACCUUAUAAUU
SARS-CoV-2-conserved-unstructured-17	25569-25582	0.825796	0.648774	UUCCAAAAUCAUAA
SARS-CoV-2-conserved-unstructured-18	19569-19588	0.75387	0.648103	UUACAAACAAUUUGAUACUU
SARS-CoV-2-conserved-unstructured-19	22552-22565	0.773461	0.646815	UAAUAUUACAAACU
SARS-CoV-2-conserved-unstructured-20	25417-25437	0.747166	0.639634	ACAAUUGGAACUGUAACUUUG
SARS-CoV-2-conserved-unstructured-21	12195-12210	0.745911	0.634404	UUAAAAAGUUGAAGAA
SARS-CoV-2-conserved-unstructured-22	6757-6783	0.78979	0.633073	UUGUACUAAUUAUAUGCCUUAUUUCUU
SARS-CoV-2-conserved-unstructured-23	15236-15257	0.747024	0.630481	ACAACAUGUUAAAAACUGUUUA
SARS-CoV-2-conserved-unstructured-24	6225-6238	0.734349	0.628462	AUAAACCUAUUGUU
SARS-CoV-2-conserved-unstructured-25	9578-9598	0.825697	0.627812	UAUUCUGUUAUUUACUUGUAC
SARS-CoV-2-conserved-unstructured-26	21649-21662	0.820527	0.626762	UAAUUCUUUCACAC
SARS-CoV-2-conserved-unstructured-27	23985-23998	0.799754	0.625346	AUCCAUCAAAACCA
SARS-CoV-2-conserved-unstructured-28	7161-7174	0.774316	0.62504	ACACCUAUCCUUCU
SARS-CoV-2-conserved-unstructured-29	6010-6029	0.738649	0.624805	ACCAAACCAACCAUAUCCAA
SARS-CoV-2-conserved-unstructured-30	6515-6529	0.81084	0.624018	UUAAAAAUUACAGAA
SARS-CoV-2-conserved-unstructured-31	18219-18232	0.749532	0.623586	UAAAAUGAAUUAUC
SARS-CoV-2-conserved-unstructured-32	11659-11681	0.819159	0.623162	UUUACUCAACCGCUACUUUAGAC
SARS-CoV-2-conserved-unstructured-33	24778-24797	0.771131	0.622469	AAAGAACUUCACAACUGCUC
SARS-CoV-2-conserved-unstructured-34	21669-21683	0.788644	0.621176	UUUAUUACCCUGACA
SARS-CoV-2-conserved-unstructured-35	6105-6122	0.777239	0.618997	AUAAGAAACCUGCUUCAA
SARS-CoV-2-conserved-unstructured-36	28436-28452	0.80827	0.618884	GCUCUCACUCAACAUGG

SARS-CoV-2-conserved-unstructured-37	16361-16375	0.772562	0.618808	UCUUGUCUGUUAAUC
SARS-CoV-2-conserved-unstructured-38	28148-28162	0.734308	0.616622	UUUUACAAUUAAUUG
SARS-CoV-2-conserved-unstructured-39	24165-24178	0.745439	0.616288	AAAUGAUUGCUCAA
SARS-CoV-2-conserved-unstructured-40	26429-26447	0.741331	0.615047	UUAAAAAUCUGAAUUCUUC
SARS-CoV-2-conserved-unstructured-41	6072-6086	0.727041	0.614595	AAUUUGCUGAUGAUU
SARS-CoV-2-conserved-unstructured-42	1304-1319	0.821877	0.614536	GAGAAUUUGACUAAAG
SARS-CoV-2-conserved-unstructured-43	1918-1933	0.751566	0.614279	UCUUGAAACUGCUCAA
SARS-CoV-2-conserved-unstructured-44	27361-27375	0.740696	0.61226	GAAGAGCAACCAAUG
SARS-CoV-2-conserved-unstructured-45	28853-28866	0.770189	0.611938	UCAAGAAAUUCAAC
SARS-CoV-2-conserved-unstructured-46	14899-14913	0.741656	0.610443	UGUAUUAAUGCUAAC
SARS-CoV-2-conserved-unstructured-47	19260-19273	0.760965	0.609632	UAACCUUAACUUGC
SARS-CoV-2-conserved-unstructured-48	11724-11740	0.729639	0.607945	UUAGAUAUAUGAAUUCA
SARS-CoV-2-conserved-unstructured-49	29008-29023	0.779731	0.607881	UGUCACUAAGAAAUCU
SARS-CoV-2-conserved-unstructured-50	11537-11554	0.771241	0.607484	AUUGUUUUUAUGUGUGUU
SARS-CoV-2-conserved-unstructured-51	11628-11645	0.85037	0.606842	AUUUUUGUACUUGUUACU
SARS-CoV-2-conserved-unstructured-52	18681-18694	0.747334	0.605637	CACAUGCUUUUCCA
SARS-CoV-2-conserved-unstructured-53	7366-7384	0.723602	0.605523	AAUAAUUAAUCUUGUACAA
SARS-CoV-2-conserved-unstructured-54	1031-1047	0.727712	0.605352	GAAAUUAAAUUGGCAAA
SARS-CoV-2-conserved-unstructured-55	14367-14380	0.714434	0.604803	UUGUGCAAACUUUA
SARS-CoV-2-conserved-unstructured-56	3797-3816	0.741669	0.60258	UUUGAUAAAAAUCUCUAUGA
SARS-CoV-2-conserved-unstructured-57	22281-22296	0.741833	0.600606	CUUUACUUGCUUUACA
SARS-CoV-2-conserved-unstructured-58	16038-16053	0.780044	0.600438	UUACCCACUUACUAAA

Table 3: **SARS-CoV-2-conserved-unstructured.** Top unstructured regions (ranked by minimum unpaired probability over the interval, stretch of at least 15 nt) that overlap with conserved intervals from SARS-CoV-2 for at least 15 nt at a 97% sequence conservation cutoff. Sequence intervals are relative to the reference genome NC_045512.2.

Name	Sequence	Secondary structure dot-bracket
5´ UTR	AUUAAAGGUUUAUACCUUCCCAGGUAACAAACCAACCAACUUUC GAUCUCUUGUAGAUCUGUUCUCUAAACGAACUUUAAAAUCUGUG UGGCUGUCACUCGGCUGCAUGCUUAGUGCACUCACGCAGUAUA AUUAAUAACUAAUUACUGUCGUUGACAGGACACGAGUAACUCGU CUAUCUUCUGCAGGCUGCUUACGGUUUCGUCCGUGUUGCAGC CGAUCAUCAGCACAUCUAGGUUUCGUCCGGGUGUGACCGAAAG GUAAGAUGGAGAGCCUUGUCCCUGGUUUCAACGAGAAAACACAC GUCCAACUCAGUUUGCCUGUUUUACAGGUUCGCGACGUGCUCG UACGUGGCUUUGGAGACUCCGUGGAGGAGGUCUUAUCAGAGGC ACGUCAACAUCUUAAAGAUGGCACUUGUGGCUUAGUAGAAGUUG AAAAAGGCGUUUUGCC	.....(((((.(((((....)))))..)))))..........(((((.....)))))..((((.....))))........((((((((.((.((((.(((.....))).)))))).)))))))(..(((((.....))))))...(((((((((((..(((((...(((.(((((((((((.((((((.(((((.....)))))..)))))))).....)))(((((((.((.....)))))))))((....)))))))))))))).)))))..))))...)))))))).....((((((.....((..((((((...))))))..))))))))(....(((((.(((((((((((((.....)))) .))))..)))))).)))))...(((((...))))))).....(((((.....)))))).....(((.....)))
Frameshifting element	GUUUUUAAACGGGUUUGCGGUGUAAGUGCAGCCCGUCUUACACCG UGCGGCACAGGCACUAGUACUGAUGUCGUAUACAGGGCUUUUG	.....((((((((((...[[[[[[[)))))))))))((((((((.....))).)))))...]].]]]]]....
3´ UTR	GACCACACAAGGCAGAUGGGCUAUAUAAACGUUUUCGCUUUUCCG UUUACGAUAUAUAGUCUACUCUUGUGCAGAAUGAAUUCUCGUAACU ACAUAGCACAAGUAGAUGUAGUUAACUUUAAUCUCACAUAGCAAUC UUUAAUCAGUGUGUAACAUUAGGGAGGACUUGAAAGAGCCACCACA UUUUCACCGAGGCCACGCGGAGUACGAUCGAGUGUACAGUGAACA AUGCUAGGGAGAGCUGCCUAUAUGGAAGAGCCCUAAUGUGUAAAA UUAAUUUUAGUAGUGCUAUCCCCAUGUGAUUUUAAUAGCUUCUUA GGAGAAUGACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA	.....((((((((((..(((...((.....))...))).)))))))))))..(((((((.....{{{{{{.[[[[[[[[[.)))))))...]]]]]]]]]...((((..(((((((((((..((..(((.(((.....(((((((((.((.(((....)))))....(((((((((.. ..((..((((.....))..))...))...)))))).)))).((((.... ....))))).....)))))))))).....))).)))..))...)))))).....)))))..))))).....}}}}}}..... .....

Table 4: Sequences and secondary structure dot-brackets for key structured genome regions of SARS-CoV-2.

References

1. Fehr, A. R.; Perlman, S., Coronaviruses: an overview of their replication and pathogenesis. *Methods Mol. Biol.* **2015,** *1282*, 1-23.
2. Connelly, C. M.; Moon, M. H.; Schneekloth, J. S., Jr., The Emerging Role of RNA as a Therapeutic Target for Small Molecules. *Cell Chem Biol* **2016,** *23* (9), 1077-1090.
3. Spurgers, K. B.; Sharkey, C. M.; Warfield, K. L.; Bavari, S., Oligonucleotide antiviral therapeutics: antisense and RNA interference for highly pathogenic RNA viruses. *Antiviral Res.* **2008,** *78* (1), 26-36.
4. Bennett, C. F.; Krainer, A. R.; Cleveland, D. W., Antisense Oligonucleotide Therapies for Neurodegenerative Diseases. *Annu. Rev. Neurosci.* **2019,** *42*, 385-406.
5. Yang, D.; Leibowitz, J. L., The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res.* **2015,** *206*, 120-133.
6. van den Born, E.; Posthuma, C. C.; Gultyaev, A. P.; Snijder, E. J., Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region. *J. Virol.* **2005,** *79* (10), 6312-6324.
7. Plant, E. P.; Dinman, J. D., The role of programmed-1 ribosomal frameshifting in coronavirus propagation. *Front. Biosci.* **2008,** *13*, 4873-4881.
8. Stammler, S. N.; Cao, S.; Chen, S.-J.; Giedroc, D. P., A conserved RNA pseudoknot in a putative molecular switch domain of the 3'-untranslated region of coronaviruses is only marginally stable. *RNA* **2011,** *17* (9), 1747-1759.
9. Chen, S.-C.; Olsthoorn, R. C. L., Group-specific structural features of the 5'-proximal sequences of coronavirus genomic RNAs. *Virology* **2010,** *401* (1), 29-41.
10. Ceraolo, C.; Giorgi, F. M., Genomic variance of the 2019-nCoV coronavirus. *J. Med. Virol.* **2020,** *92* (5), 522-528.
11. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Hu, Y.; Song, Z.-G.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; Yuan, M. L.; Zhang, Y.-L.; Dai, F.-H.; Liu, Y.; Wang, Q.-M.; Zheng, J.-J.; Xu, L.; Holmes, E. C.; Zhang, Y.-Z., NC_045512. Nucleotide [Internet]. Bethesda (MD): National Library of Medicine (US), N. C. f. B. I., Ed. 2020.
12. Viruses, C. S. G. o. T. I. C. o. T. o.; Coronaviridae Study Group of the International Committee on Taxonomy of, V., The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* **2020,** *5* (4), 536-544.
13. Madhugiri, R.; Fricke, M.; Marz, M.; Ziebuhr, J., RNA structure analysis of alphacoronavirus terminal genome regions. *Virus Res.* **2014,** *194*, 76-89.
14. Kalvari, I.; Argasinska, J.; Quinones-Olvera, N.; Nawrocki, E. P.; Rivas, E.; Eddy, S. R.; Bateman, A.; Finn, R. D.; Petrov, A. I., Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic Acids Res.* **2018,** *46* (D1), D335-D342.
15. Nawrocki, E. P.; Eddy, S. R., Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **2013,** *29* (22), 2933-5.
16. Yang, D.; Liu, P.; Wudeck, E. V.; Giedroc, D. P.; Leibowitz, J. L., SHAPE analysis of the RNA secondary structure of the Mouse Hepatitis Virus 5' untranslated region and N-terminal nsp1 coding sequences. *Virology* **2015,** *475*, 15-27.
17. Elbe, S.; Buckland-Merrett, G., Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **2017,** *1* (1), 33-46.
18. Do, C. B.; Woods, D. A.; Batzoglou, S., CONTRAfold: RNA secondary structure prediction without physics-based models. *Bioinformatics* **2006,** *22* (14), e90-8.
19. Hamada, M.; Sato, K.; Asai, K., Prediction of RNA secondary structure by maximizing pseudo-expected accuracy. *BMC Bioinformatics* **2010,** *11*, 586.
20. Herold, J.; Siddell, S. G., An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res* **1993,** *21* (25), 5838-42.

21. Gruber, A. R.; Findeiß, S.; Washietl, S.; Hofacker, I. L.; Stadler, P. F., RNAz 2.0: improved noncoding RNA detection. *Pac. Symp. Biocomput.* **2010**, 69-79.
22. Rivas, E.; Clements, J.; Eddy, S. R., Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* **2020**.
23. 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* (3), 790-803.
24. 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* (5), 763-780.
25. Kelly, J. A.; Dinman, J. D., Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2. *bioRxiv* **2020**.
26. Goebel, S. J.; Hsue, B.; Dombrowski, T. F.; Masters, P. S., Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J. Virol.* **2004,** *78* (2), 669-682.
27. Goebel, S. J.; Miller, T. B.; Bennett, C. J.; Bernard, K. A.; Masters, P. S., A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *J. Virol.* **2007,** *81* (3), 1274-1287.
28. Robertson, M. P.; Igel, H.; Baertsch, R.; Haussler, D.; Ares, M., Jr.; Scott, W. G., The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol.* **2005,** *3* (1), e5.
29. Jonassen, C. M.; Jonassen, T. O.; Grinde, B., A common RNA motif in the 3' end of the genomes of astroviruses, avian infectious bronchitis virus and an equine rhinovirus. *J. Gen. Virol.* **1998,** *79 (Pt 4)*, 715-718.
30. Guan, B.-J.; Su, Y.-P.; Wu, H.-Y.; Brian, D. A., Genetic evidence of a long-range RNA-RNA interaction between the genomic 5' untranslated region and the nonstructural protein 1 coding region in murine and bovine coronaviruses. *J. Virol.* **2012,** *86* (8), 4631-4643.
31. Kappel, K.; Zhang, K.; Su, Z.; Kladwang, W.; Li, S.; Pintilie, G.; Topkar, V. V.; Rangan, R.; Zheludev, I. N.; Watkins, A. M.; Yesselman, J. D.; Chiu, W.; Das, R., Ribosolve: Rapid determination of three-dimensional RNA-only structures. *bioRxiv* **2019**.
32. Zhang, K.; Li, S.; Kappel, K.; Pintilie, G.; Su, Z.; Mou, T.-C.; Schmid, M. F.; Das, R.; Chiu, W., Cryo-EM structure of a 40 kDa SAM-IV riboswitch RNA at 3.7 Å resolution. *Nat. Commun.* **2019,** *10* (1), 5511.
33. Edgar, R. C., MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **2004,** *32* (5), 1792-1797.
34. Li, W.; Jaroszewski, L.; Godzik, A., Clustering of highly homologous sequences to reduce the size of large protein databases. *Bioinformatics* **2001,** *17* (3), 282-283.
35. Sievers, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G., Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **2011,** *7*, 539.
36. Katoh, K.; Frith, M. C., Adding unaligned sequences into an existing alignment using MAFFT and LAST. *Bioinformatics* **2012,** *28* (23), 3144-3146.
37. Washietl, S.; Hofacker, I. L., Consensus folding of aligned sequences as a new measure for the detection of functional RNAs by comparative genomics. *J Mol Biol* **2004,** *342* (1), 19-30.
38. Rivas, E.; Clements, J.; Eddy, S. R., A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **2017,** *14* (1), 45-48.
39. Cock, P. J.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J., Biopython: freely available Python tools for computational molecular biology and bioinformatics. *Bioinformatics* **2009,** *25* (11), 1422-3.
40. Bernhart, S. H.; Hofacker, I. L.; Stadler, P. F., Local RNA base pairing probabilities in large sequences. *Bioinformatics* **2006,** *22* (5), 614-5.
41. Kearse, M.; Moir, R.; Wilson, A.; Stones-Havas, S.; Cheung, M.; Sturrock, S.; Buxton, S.; Cooper, A.; Markowitz, S.; Duran, C.; Thierer, T.; Ashton, B.; Meintjes, P.; Drummond, A., Geneious

Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **2012,** *28* (12), 1647-1649.