

Towards elucidating the structure, function, and druggability of coronavirus cis-acting RNA motifs

ABSTRACT

Coronaviruses are the causative agents of mild to severe respiratory and intestinal infections in humans. They are the largest RNA viruses, which genomes and encoded RNAs are known to fold into the highly-order structures that play essential roles in the viral replication and infectivity cycle. The recent outbreaks of new pathogenic coronaviruses steered researchers' attention into the possibility of targeting their RNAs directly with novel RNA-specific drugs and therapeutic strategies. In this manuscript, we highlight the recent biochemical and biophysical methodological advancements that yielded more in-depth insight into the structural and functional composition of coronaviruses *cis*-acting RNA motifs. We discuss the complexity of these RNA regulatory elements, their intermolecular interactions, post-transcriptional regulation, and their potential as druggable targets. We also indicate the location and function of unstructured and highly-conserved regions in coronaviruses RNA genomes representing viable aims for antisense oligonucleotide or CRISPR-based antiviral strategies.

INTRODUCTION

Coronaviruses are positive-sense, single-stranded (ss) RNA viruses that belong to the family Coronaviridae, further divided into four genera, i.e., alpha, beta, gamma, and delta. Each genus includes closely related viruses collected into specific lineages or groups [1,2]. Recent cross-species transmission events and changes in virus tropism have triggered the emergence of new pathogenic coronaviruses. Severe acute respiratory syndrome coronavirus (SARS-CoV), middle east respiratory syndrome coronavirus (MERS-CoV), and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), are the prominent examples of highly infectious human coronaviruses, all belonging to the beta genus [2]. Apart from these, HCoV-OC43, and HKU1, which also belong to the beta genus, as well as HCoV-229E and NL63, classified under the alpha genus, cause mild respiratory infections in humans [3].

While most of the approved drugs act on protein targets, viral RNAs are increasingly recognized as key regulatory players of molecular processes, that can be specifically and effectively targeted with drugs or therapies [4]. With their well-defined structure, many RNA folds provide potentially unique interaction sites for selective binding of bioactive molecules that can disturb the RNAs' function. Ligands binding to RNA, whether it involves a small molecule chemotype or antisense oligonucleotide (ASO), can not only affect structural stability or invoke RNA conformational change, but also it can disrupt the intermolecular interactions, and thereby block processes essential for the viral replication [5]. The CRISPR/Cas-based system yields also new therapeutic tools for targeting viral genomes and virus-encoded RNAs [6]. Here, the knowledge of highly conserved and unstructured regions across viral RNAs is essential for their effectiveness and contribution to the antiviral armamentarium.

Recent advancements in the field of RNA biology, including, but not limited to the development of diverse sets of chemical approaches for the RNA structure probing [7], direct RNA sequencing [8], computational predictions [9], and the expansion of next-generation sequencing tools [10,11], yielded more comprehensive insight into the secondary and tertiary structure adopted by viral RNA motifs. This knowledge will undoubtedly encourage a more extensive search for RNA-based therapeutics.

In this manuscript, we provide a broad overview of *cis*-acting RNA motifs that can be found within the RNA genomes of human coronaviruses. We discuss their functionality, structure, post-transcriptional modifications, and their potential as therapeutic targets. Our focus will be on the most recent works that have shed significant insight into the RNA structural conformation of the newly

Joanna Sztuba-Solinska[✉],
Devadatta Gosavi

Department of Biological Sciences, Auburn University, 120 W. Samford Ave, Rouse Life Sciences Building, Auburn, AL 36849, United States

https://doi.org/10.18388/pb.2019_

[✉]autor korespondujący: jzs0165@auburn.edu

Key words: coronaviruses, RNA structure, *cis*-acting RNA motifs, epitranscriptomic modifications, RNA targeting, therapy

Abbreviations: CRISPR/Cas - clustered regularly interspaced short palindromic repeats/CRISPR-associated protein 9; DRS - Direct RNA sequencing; FSE - frameshift element; HCoV - human coronavirus; ORF - open reading frame; RAP-MS - RNA antisense purification and mass spectrometry; SARS - severe acute respiratory syndrome coronavirus; SHAPE-MaP - Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling; TRS - transcription-regulating sequences; UTR - untranslated region

emerged SARS-CoV-2. We also reference previous key studies related to other human coronaviruses' RNA structure and function, to provide a prelude to recent developments.

IDENTIFYING THE CIS-ACTING RNA MOTIFS IN HUMAN CORONAVIRUSES GENOMES.

The coupling of an increasingly diverse set of tools and techniques for structural analysis of viral RNA has provided significant insight into their functionality, more in-depth characterization of their folding pathways, and the identification of novel cis-acting RNA motifs that can be potentially targeted with antivirals. Initially, the majority of coronavirus RNA structural studies have been performed based on *in vitro* systems. Thermodynamic and kinetic studies using nuclear magnetic resonance (NMR) spectroscopy provided insight into the complex folding pathways of the 5' untranslated region (UTR) and frameshift element (FSE) [12]. Homo and heteronuclear two-dimensional NMR have confirmed the base pairing interactions involved in the FSE pseudoknot formation in SARS [13]. NMR has also been used to determine the solution structure of highly conserved stem-loop 2 (SL 2) present in the 5' UTR of SARS-CoV [14].

In vitro biochemical structure probing techniques further shed light on the structures of functional RNA motifs at the single-nucleotide level. Several RNA cleaving enzymes and chemical probes have been employed to attack and modify the RNA bases, sugar, and backbone, to reveal the base pairing status of the nucleotides. A set of RNases, including A, T1, and V1 that cleave at single-stranded (ss) pyrimidines, ss guanines, and double-stranded (ds) RNA, respectively, have been used to examine the SARS-CoV FSE, confirming the computational predictions and verifying the formation of all three stems [13]. Also, dimethyl sulfide (DMS), which modifies unpaired adenosines and cytidines, has been used to study the 5'-terminal ~100 nt of the HCoV-229E and HCoV-NL63 genome indicating transcription-regulating sequences (TRSs) localization within ss region [15].

However, RNA structural studies performed *in vitro* are often deficient in the direct relevance to what happens inside the living cell, as the solution conditions lack many components that influence RNA folding and function [16]. Since RNA structure studies' ultimate goal is to understand how RNA behaves in the native conditions, the majority of the methods developed to study RNA *in vivo* are structure probing. Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) is a probing technique that relies on the application of cell-permeable electrophilic reagent, i.e., 2-methylnicotinic acid imidazolide (NAI), 1-methyl-7-nitroisatoicanhydride (1M7) that modify the 2' OH group of ribose in accessible RNA nucleotides, leaving the base paired or inaccessible residues largely unmodified [17]. The modification of the target RNA leads to the formation of 2'-O-adducts, which during the reverse transcription performed in the presence of manganese ions, invoke mutation relative to the sequence complementary to the RNA being copied. The obtained cDNA products are directed to the stepwise amplification, followed by next-generation sequencing and bioinformatics analysis. This analysis reveals the Shannon entropy and SHAPE-MaP

reactivity values, which can be further utilized in the Shape-Knots pipeline for predicting pseudoknots. Most recently, SHAPE-MaP has been used to characterize the full-length SARS-CoV-2 RNA genome [5]. Also, *in vivo* click selective 2'-hydroxyl acylation and profiling experiment (icSHAPE) has been performed to study the SARS-CoV-2 RNA secondary structure in infected cells [18]. icSHAPE provides accurate predictions of RNA secondary structure *in vivo* by combining SHAPE chemistry with click chemistry for enhancing isolation and reduced signal-to-noise ratio. Secondary RNA structures are modified by the addition of 2-methylnicotinic acid imidazolide probe, termed NAI-N3, which reacts with ss nucleotides and modifies the 2' OH group. DMS mutational profiling and sequencing (DMS-MaPseq) is another biochemical probing technique that has been used for the determination of the SARS-CoV-2 RNA secondary structure. DMS-MaPseq relies on the use of DMS and a thermostable group II intron level transcriptase that during reverse transcription of target RNA reads through the DMS modified adenosines and cytosines and records them as mismatches in the cDNA product. DMS-MaPseq is known to provide high-quality RNA structure data with a high signal-to-noise ratio [19].

The recent development of Direct RNA sequencing (DRS) using MinION nanopore sequencer has found application in the deconvolution of the primary sequence of SARS CoV-2 transcriptome. DRS enables long-read sequencing, which is particularly helpful for the analysis of long nested coronavirus transcripts. As DRS detects RNA directly, it also provides useful information regarding the RNA epitranscriptomic modifications [20]. In DRS, a single RNA molecule is ratcheted through a protein pore fixed in a synthetic membrane using a molecular motor [9]. Based on the chemical composition of nucleobases, the passage of RNA molecules through the narrowest section of the pore modifies the movement of ions across the membrane. If the nucleotides are modified, then there is an apparent change in the current intensity, and in the time, RNA molecule resides inside the pore (dwell time). These parameters are useful for the identification of not only the primary sequence of RNA but also its post-transcriptional modifications [9]. Intriguingly, the DRS has been recently coupled with SHAPE biochemistry in the method referred to as PORE-cupine, for chemical utilized probing interrogated using nanopores. PORE-cupine identifies RNA ss nucleotides by detecting current changes induced by structure modifications, opening the possibility of the fast and direct assay for RNA structures and dynamics genome-wide [21].

In silico RNA structure predictions have also been widely used to understand the coronaviruses RNA structure and functions. Homology modeling is the most precise computational method for creating reliable RNA structural models and is used when only the sequence of target RNA is known, and the three-dimensional (3D) structure is yet to be determined. In this method previously determined 3D structure of RNA, whose sequence is similar to the target RNA, is utilized for arranging the backbone of the RNA in the query. Further, the target and template sequences are aligned, and the 3D structure of RNA is generated. Several methods have been proposed for 3D modeling of an RNA

structure, with FARNA and FARFAR being most widely used. FARNA involves assembling of RNA 3D structure from short linear fragments by using a knowledge-based energy function, which considers the backbone and side-chain conformations, as well as base pairing and base stacking interactions. FARFAR is an extension of FARNA and utilizes full-atom refinement for optimizing the RNA structures generated by FARNA [22]. Recently, Rosetta's fragment assembly of RNA with full-atom refinement and FARFAR2 algorithm have been used to generate the *de novo* models of SARS-CoV-2 *cis*-acting RNA elements, i.e., stem-loops within 5' UTR and 3' UTRs [23]. SARS-CoV-2 genome structure has also been analyzed using an RNA motif pipeline known as ScanFold [24]. The ScanFold generates unique two-dimensional (2D) models for highly-structured and likely functional motifs based on the minimum free-energy and partition function calculations. This method is not only valuable for functional RNA structured motif discovery and mapping the general RNA folding landscape, but also for identifying structures likely to be available for targeting via small molecules. Other computational algorithms, including RNAz and Contrafold, have provided further insights into the secondary structure of SARS-CoV-2 [25]. RNAz predicts structured regions that are more thermodynamically stable than expected by comparison to random sequences of the same length and sequence composition and assesses regions by the support of compensatory and consistent mutations in the sequence alignment. Contrafold, on the other hand, predicts RNA secondary structures without physics-based models, and instead uses learned parameters based on known structures. Alike *in vitro* techniques, the drawback of *in silico* generated models is that they cannot account for the physiological conditions that influence RNA fold and behavior inside the cells.

Not only viral RNA secondary structure, but also RNA-mediated interactions with viral and cellular effectors, i.e., other RNAs and proteins, can be recognized as valuable targets for the development of antiviral therapeutics. Here, the disruption of an identified interaction would be expected to lead to interference with the viral infectivity cycle. The recently developed tool, termed PrismNet (Protein RNA Interaction by Structure-informed Modeling using deep neural Network), has been utilized for predicting RNA binding proteins (RBP) and their binding sites on SARS-CoV-2 RNA genome [26]. The principle behind PrismNet is that it aids in constructing and training a neural network for modeling the interaction between an RBP and its RNA target by incorporating data from *in vivo* RBP binding assays and RNA structural analysis obtained from similar cellular conditions. The application of PrismNet has led to the identification of 40 host proteins that bind to the 5' UTR and 43 proteins that bind to 3' UTR in the SARS-CoV-2 genome [18]. These include proteins involved in stress granule formation, i.e., TIA1, ELAVL1, and autoimmune disorders, i.e., TROVE2. Also, RNA antisense purification and mass spectrometry (RAP-MS) has found an application to obtain an unbiased and quantitative picture of the human proteome that directly binds the SARS-CoV-2 RNA in infected human cells [27]. RAP-MS relies on ultraviolet crosslinking of the interacting target RNA and proteins, followed by the target RNA's affinity capture and mass spectrometry analysis of

the RBPs. The application of highly denaturing purification conditions allows identifying only the directly interacting proteins with high specificity [28]. The RAP-MS analysis has identified numerous host factors, including regulators of RNA metabolism, translation, and host defense pathways, as required for SARS-CoV-2 replication.

THE *CIS*-ACTING RNA MOTIFS IDENTIFIED WITHIN THE HUMAN CORONAVIRUSES GENOME

Coronaviruses have the largest genomes known among RNA viruses, ranging from 27 to 32 kb, and share similar structural organization (Fig. 1) They are 5' capped and 3' polyadenylated and enclose multiple open reading frames (ORFs). The 5' coding region comprises of RNA-dependent RNA polymerase (RDRP) gene, which extends over two-thirds of the genome, and contains two overlapping ORFs, while the downstream region codes for the structural proteins, and non-essential accessory proteins [29]. The coronaviral RNA folds back on itself, resulting in complex secondary and tertiary structures, known as the *cis*-acting RNA motifs. These *cis*-acting motifs have been shown to comprise a series of conserved stem-loops located in the 5' UTR, a pseudoknot enclosed within the FSE, and mutually exclusive structures, including another pseudoknot or a stem-loop formed within the 3' UTR [30]. Furthermore, the 5' UTR encloses the TRS leader sequence (TRS-L), which is essential for facilitating the discontinuous transcription characteristic to coronaviruses [31]. These RNA structural elements are functionally necessary for the overall genome stability, RNA-RNA interactions, and for the binding of viral and cellular proteins during RNA replication, transcription, and translation [30,32].

The organization of 5' untranslated region (5' UTR). The coronaviruses' very 5' 350 nucleotides fold into a set of highly conserved structural repeats, i.e., stem-loops, that have been shown to mediate interactions with the membrane (M) and nucleocapsid (N) proteins to facilitate the effective packaging of viral RNA [23]. Reverse genetic studies have shown that mutations that disrupt these structures impede the progeny virion production [33]. The 5' UTR also mediates translation initiation through the canonical cap-dependent ribosomal entry mechanism [23]. This process requires the establishment of long-range interaction between the 5' and 3' UTRs, leading to the genome circularization, which is critical not only for the viral transcription but also RNA replication [30]. The 5' UTRs also contains the TRS-L, which includes a conserved core sequence (5'-ACGAAC-3'). This core sequence is also found upstream of each ORF and is referred to as body TRSs (TRS-B) [19]. The RDRP has been proposed to pause after the TRS-B of each gene during the negative-sense RNA synthesis, subsequently switching to the TRS-L, and thus, adding a common leader (L) sequence to each subgenomic (sg) RNA. This mechanism, referred to as the discontinuous transcription, leads to the fusion of L-B sequences based on the complementarity of nascent negative-sense RNA with positive-sense TRS-L. As a result, the 5' nested set of negative-sense sgrNAs are formed, which are further used to synthesize of the 3' nested set of positive-sense sgrNAs [20]. These sgrNAs encode virulence

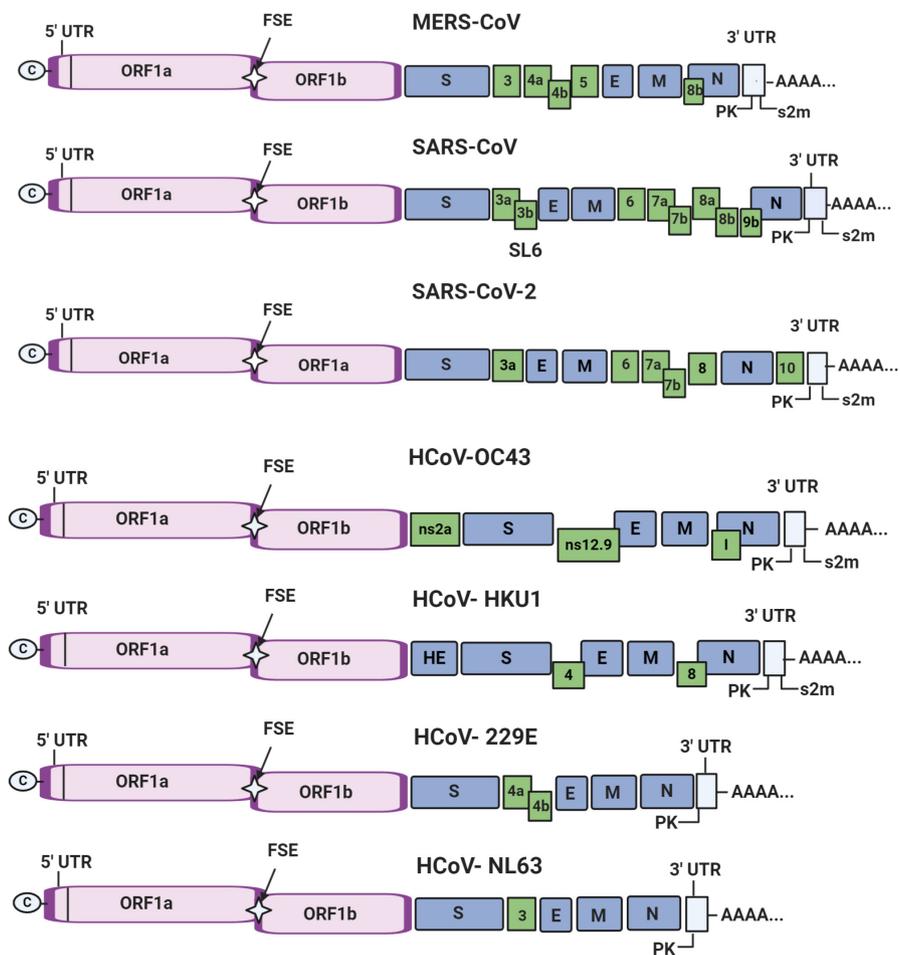


Figure 1. Schematic representation of the genomic structure of human coronaviruses (HCoV).

The 5' cap structure and the 3' poly A tail (AAAA...) are indicated at the 5' and 3' untranslated regions (UTRs), respectively. The open reading frame 1a (ORF 1a) and ORF 1b are shown in purple. The star between the overlapping ORF's indicates the Frameshifting Element (FSE). The genes encoding structural proteins, i.e., spike (S), envelope (E), membrane (M), nucleocapsid (N) and hemagglutinin-esterase (HE) are shown as blue boxes. The genes encoding accessory proteins are shown in green boxes. Pseudoknot (PK) and stem-II like motif (s2m) are indicated at the 3' UTR.

factors and have been shown to influence the host immune response [34].

Recent applications of the DMS-MAPseq [19] SHAPE-MaP probing [6] and RNA structure modeling FARFAR algorithm [23] have provided in-depth insight into the structural conformation of the 5' UTRs in human coronaviruses. The SARS-CoV-2 5' UTR has been shown to include five stem-loops, termed as SL 1 – 5 (Fig. 2) [25]. The SL 1 adopts a bipartite stem, in which two helices are separated by bulged adenosine at the 5' side and the AC bulge on the 3' side. In general, the SL 1 is AU rich with its boundary marked by two consecutive GC base pairs. The SL 2 includes a U-turn motif and a penta-loop 5'-(C/U)UUG(U/C)-3', which stacks on its 5-nucleotide stem [19]. *In vitro* studies have indicated that mutations in SL 2 penta-loop lead to the disruption of sgRNA synthesis [19]. The SL 3 contains the TRS-L sequence, 5'-ACGAAC-3' that localizes to the 3' side of its stem. Structural probing analysis of SL 3 has revealed that it includes residues of medium reactivity towards probing reagent, which suggests that SL 3 transitions between different conformation [19]. Interestingly, SHAPE-MaP probing of that region in MERS-CoV predicts the absence of SL 3 [35]. The

SL 4 is a bulged bipartite stem-loop divided into two stems, SL 4a and SL 4b. This structure includes three non-canonical GU base pairs and a short upstream open reading frame (uORF), whose AUG codon localizes to the stem's apical loop [19]. Reverse genetics and *in vitro* studies have shown that mutations in uORF lead to the moderate reduction of viral RNA replication [2]. The SL 5 is a well-established domain containing one main stem, that connects three stem-loops: SL 5A, SL 5B, SL 5C. The SL 5A and 5B both contain 5'-UUCGU-3' penta-loop, while SL 5C is closed by the apical GNRA tetra-loop [36]. This tetra-loop has moderate reactivity towards SHAPE reagents in most recent probing experiments [26, 37], likely due to the extensive stacking interactions between adenosines [38]. The 3' side of the main stem includes the AUG start codon, just downstream of SL 5C. Overall, SHAPE-MaP probing accompanied by Shannon entropy analysis [37] have indicated that the 5' UTR of SARS-CoV-2 displays low Shannon and low SHAPE reactivity, which has been previously proposed to be characteristic of well-defined functional regions in ZIKA virus (ZIKV) [10], Dengue virus (DENV) [10], Hepatitis C virus (HCV) [11], and Human immunodeficiency virus (HIV 1) [39] genomes.

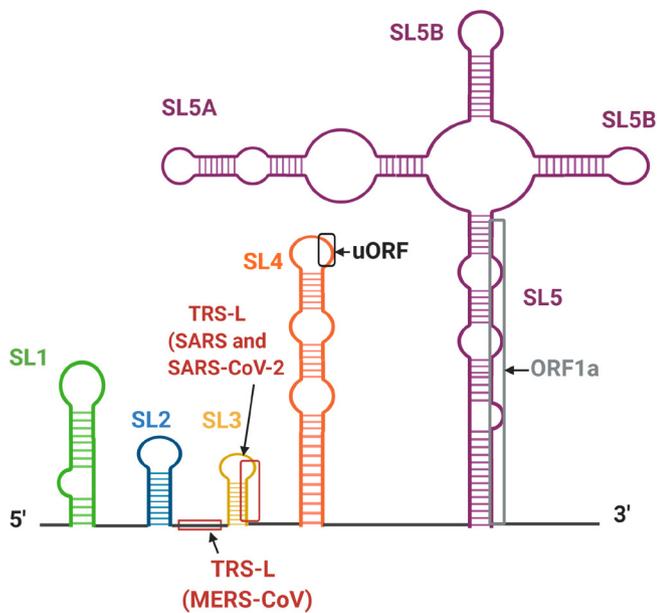


Figure 2. RNA secondary structure of the 5' UTR in SARS, SARS-CoV-2 and MERS.

Five stem-loops, i.e., SL 1 (green), SL 2 (blue), SL 3 (yellow), SL 4 (orange), and SL 5 (purple) are indicated. The position of leader sequence within the TRS-L in SARS and SARS-CoV-2 is enclosed within SL 3, and indicated with red rectangle. The TRS-L in MERS-CoV is located between SL 2 and SL 3, and marked also with red rectangle. The upstream open reading frame (uORF) within SL 4 is marked with a black rectangle. The beginning of ORF 1a structure is indicated by a grey box on SL 5.

These regions are likely to be engaged in the mechanistic and structural aspects of the viral RNA function.

The 5' UTRs structure predictions for other human beta-coronaviruses, i.e., HCoV-OC43, HCoV-NL63, and HCoV-229E, have mostly been based on the comparative analyses and sequence homology modeling [2], followed by side-directed mutagenesis experiments [30]. It has been shown that the SL 1 and SL 2 are largely conserved, with SL 2 showing the highest sequence conservation [2], while SL 5 displaying some degree of sequence and structure variability. For alpha-coronaviruses, the reverse genetics studies have confirmed the functional importance of SL 1 and SL 2, as single-nucleotide substitutions predicted to destabilize these structures abolished viral RNA synthesis [40]. Interestingly, HCoV-229E SL 2 can be replaced with that of SARS-CoV SL 2, providing experimental support that some RNA structural elements in the coronavirus 5' UTR display functional conservation. The authors concluded that the SL 1 and SL 2 secondary structure is more important for viral replication than preserving a specific nucleotide sequence [40]. Also, multiple sequence alignments suggest that the TRS-L core sequence, SL 4, and SL 5, despite their poorly defined structures, are conserved among alpha-coronaviruses [4]. Recent icSHAPE-based analysis of the MERS-CoV 5' UTR indicated that it contains almost identical stem-loops, even though the primary sequence has a sequence similarity of only ~46% with lineage B, represented by SARS-CoV [26]. Strikingly, despite having similar levels of sequence similarity to the consensus reported for beta-coronaviruses, the lineage A of beta-coronavirus represented by HCoV-HKU1 and alpha-coronavirus, represented by HCoV-NL63 (37.5%~

46.3% in 5'UTR) each have been proposed to form very different structures, although all include SL 1. This structural divergence suggests that these viruses' non-coding regions may be subject to different forms of regulation.

The structure of frameshifting element (FSE). The FSE is present in the first protein-coding ORF (ORF 1ab), and it includes sequences and structures that are essential for the programmed (-1) ribosomal frameshifting and translation of ORF 1ab polyprotein [6,23]. The structure of the FSE of SARS-CoV has been initially solved by NMR, which indicated the formation of a three-stemmed pseudoknot [13,25]. The prevailing mechanism is that the pseudoknot causes the ribosome to pause at the slippery sequence and backtrack by one nucleotide to release mechanical tension. Due to this shift, the ribosomes can bypass a canonical stop codon, which facilitates efficient translation. Simultaneously, the slippery sequence helps in re-pairing and continuing elongation of the polypeptide in the new translational reading frame [41].

The recent comparative structural analysis of SARS-CoV-2 RNA has indicated more in-depth inside into the composition of the FSE signal. This RNA domain has been proposed to begin with a heptameric sequence of the slippery site followed by a six-nucleotide spacer. In SARS-CoV-2, the heptameric sequence is 5'-UUUAAAC-3'; the introduction of mutations within that region ablates the ribosomal frameshifting [41]. The stimulatory region that is present at the 3' of FSE consists of three canonical stems involved in forming a pseudoknot. The stem 1 is predicted to be 11 nucleotides long, while the stem 2 and 3 are seven and eight nucleotides long, respectively, and are separated by a bulge (Fig. 3) [23,41]. SARS-CoV and SARS-CoV-2 RNAs pairwise sequence alignment has indicated that this bulge includes a single residue that distinguishes between both viruses (C for SARS-CoV and A for SARS-CoV-2), but it does not alter the overall fold of FSE [41]. However, the mutations that disrupt the stem 1 and 2 have been shown to abolish the FSE fold leading to detrimental effects on viral propagation. In contrast, the disruption of stem 3 does not completely abolish the process [41].

The recent SHAPE-MaP probing of the SARS-CoV-2 FSE indicated that this region shows low Shannon and low SHAPE values, emphasizing its structural and functional importance [37]. The application of ShapeKnots incorporating SHAPE-MaP reactivities confirmed the base pairing pattern specific for the pseudoknot formation. Interestingly, the DMS-MaPseq probing has resulted in an alternative model that did not include the pseudoknot [19]. Instead, the *in cellulo* model included Alternative Stem 1 (AS1) that forms when half of the canonical stem 1 finds an alternative pairing partner driven by 10 complementary bases upstream of the slippery site. A similar structure has been proposed by *in silico* predictions using RNAz [24] and ScanFold [23]. In SARS-CoV-2, ScanFold not only predicted the AS1 but also found that this structure is more stable relative to any other structure in the entire FSE [23]. Subsequently, the detection of RNA folding ensembles using the expectation-maximization algorithm and DMS probing data showed that this RNA region folds into at least two distinct conformations,

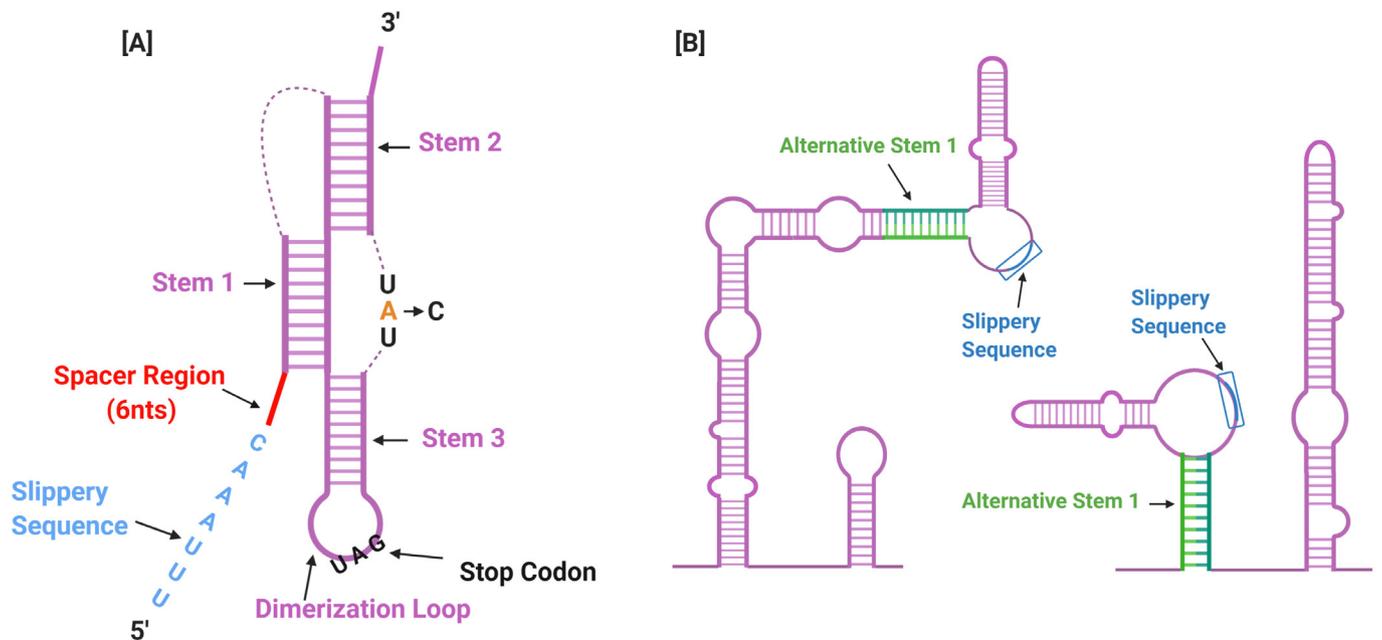


Figure 3. Secondary structure models of Frame Shift Element (FSE) and its alternative conformations.

[A] The FSE is composed of the 5' slippery site (UUUAAAC) highlighted in blue, a 6-nts spacer region indicated in red, and three-stem pseudoknot stimulatory element. A single nucleotide difference between SARS-CoV-2 and SARS-CoV (A → C) is indicated in the bulge linking stem 2 and 3. [B] Structural secondary structure model of the FSE alternative conformations derived from DMS-MaPSeq. Alternative stem is highlighted in green-blue and the slippery site boxed in blue.

both involving the formation of the Alternative Stem 1, but not a pseudoknot.

Another regulatory element referred to as the attenuator hairpin has been proposed to form in the upstream of the slippery site and can diminish -1 programmed ribosomal frameshifting. In SARS-CoV-2, this hairpin is nine nucleotides long with a G bulge and AGCU tetra-loop. This sequence in the SARS-CoV-2 reporter, significantly decreased the programmed ribosomal frameshifting [41]. In SARS-CoV, on the other hand, the attenuator hairpin contains the G bulge and UGCG tetra-loop, instead. Neither the replacement of G in the bulge nor the insertion of six nucleotides 5'-ACGACU-3' in the loop hindered the attenuation efficiency. However, the deletion of six nucleotides at the 5' half of its stem abolished the attenuation significantly. Additionally, a mutational study has indicated that the spacing between the attenuator hairpin and spacer region is critical for the attenuation activity [42]. Interestingly, the computational prediction studies performed for the FSE in HCoV-229E and HCoV-NL63 have suggested the formation of a distinct type of pseudoknot called as “elaborated pseudoknot” or “kissing stem-loop” [35]. The predicted structure consisted of stem 2 of only 5 base pairs and a large, atypical loop connecting stem 1 and stem 2. A computer-assisted analysis also predicted a third stem-loop, involved in base pairing on either side of 3' component of stem 2 and required for high-frequency frameshifting [43].

The cis-acting RNA motifs enclosed within the coding regions. The coronaviruses protein-coding regions have also been predicted to fold into cis-acting RNA motifs that can regulate critical aspects of viral replication and pathogenesis. The application of DMS-MaP probing has recently re-

sulted in the prediction of three stem-loops, SL 6, 7, 8, that lie downstream of the 5' UTR within the coding sequence of nonstructural protein 1 in SARS-CoV and SARS-CoV-2 [44]. Although, previous *in silico* model of SARS-CoV-2 has suggested the formation of three short stem-loops in place of SL 8 [28]. The recent ScanFold prediction of SARS-CoV-2 RNA has indicated the unusually strong folds existing within ORF 3a [24]. These include a stretch of 10 predicted hairpins, which do not show evidence of specific base pair conservation. The same region in SARS-CoV appears similarly structured, despite that it shows only 68% sequence similarity to SARS-CoV-2. Besides, structure predictions performed for SARS-CoV-2 have shown that TRS-B located within seven ORFs (3a, 6, 8, S, M, E, N) contain the core sequence 5'-ACGAAC-3' within stem-loops [19].

Further, the application of DMS-MaPseq has shown that TRS-B present at ORF 6 contains two internal loops and a 2-nucleotide bulge, while TRS-B at ORF 8 includes two internal loops with the core sequence partially involved in the internal loop. The RNA region coding for M protein has also been shown to fold into two small bulges and an internal loop. In contrast, the regions coding for N protein folds into three internal loops with the core sequence covering one of the loops (Fig. 4).

Previously, it has been proposed that the levels of single-strandedness within the 5' termini of given RNA associates with the relative abundance of a particular transcript [45]. The correlation of the *in vivo* icSHAPE analysis for SARS-CoV-2 TRS-Bs with the translation efficiency of individual sgrNAs, confirmed these findings, as the 5' single-strandedness of the TRS-Bs influences the relative abundance of sgrNAs, presumably because of differential

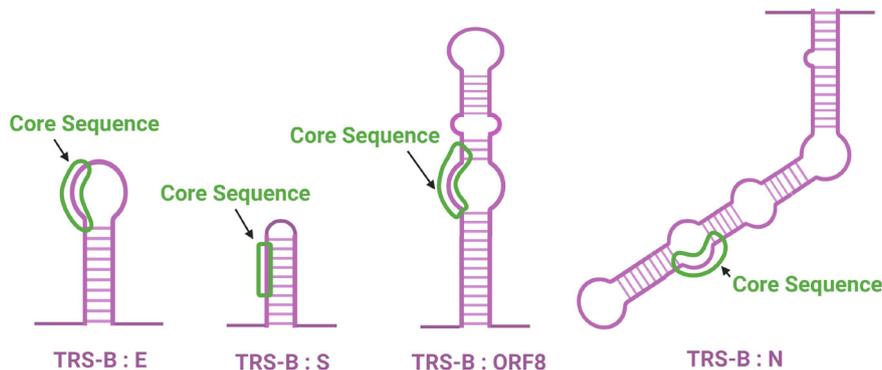


Figure 4. Secondary structure models predicted for four TRS-Bs in SARS-CoV-2 RNA.

The core sequence in each stem-loop is highlighted in green. E indicates envelope protein gene, S is for spike protein gene, ORF 8 stands for Open ReadinG Frame 8, and N indicates nucleocapsid protein gene.

impact on discontinuous transcription [18]. In general, the DMS-MaPseq analysis of the SARS-CoV-2 RNA structure has shown that over 21% of its genome includes accessible, i.e., ss regions [19]. Every ORFs, except for ORF E, have been shown to contain at least one of these accessible regions with the two longest unpaired stretches occupying ORF 1a and S. These regions may offer multiple binding sites for antisense oligonucleotide-based therapy.

The organization of 3' untranslated region (3' UTR). The human coronaviruses' 3' terminus contains pivotal domains for the regulation of viral RNA synthesis and the recruitment of host translation machinery. The 3' UTR has been shown to bind to the host translation initiation factors to hijack the cellular translational machinery for its use [46]. The 3' UTR is also involved in the 5'-3' genome circularization [25].

In SARS-CoV-2, the 3' UTR comprises of a switch-like domain involving the formation of H-type pseudoknot (PIPk), stem-loop 2-like motif (s2m) and a hypervariable region (Fig. 5) [23]. The pseudoknot involves canonical base pairs that scaffold three stems, i.e., P2, P0b, P5, with P2 stem positioned between P0b and P5, and connected with them by one nucleotide loop 1 and 3 nts loop 2. The pseudoknot

formation has been proposed to be mutually exclusive with the formation of a P0b stem, which contains an apical hexa-loop and a bulge with two adenines. Additionally, the base pairing involving loop 1 of the pseudoknot, results in the formation of P4 and P5 stems [25]. The 3' UTR pseudoknot, along with its mutually exclusive stem-loop have been suggested to regulate viral RNA synthesis, as the introduction of mutations that destabilize either of these structures, disrupts the viral replication in beta-coronaviruses [47]. Recent icSHAPE analysis has proposed the formation of a stem-loop rather than a pseudoknot in the 3'UTR of the SARS-CoV-2 genome and emphasized its overall high-level of single-strandedness. These findings highlight that *in vivo* structural information is critical for building physiologically relevant structural models [26].

The s2m is a sub-region within the hypervariable region, which resembles the ribosomal RNA loop structure. Thus, it has been proposed to bind translation initiation proteins [25,48]. The s2m is defined by two perpendicular RNA helix's axes containing one internal loop, two asymmetric bulges, and one apical 5'-GAGUA-3' penta-loop, similar to conventional GNRA tetra-loop, but with an extra bulged U [25,48]. On the other hand, the hypervariable region consists

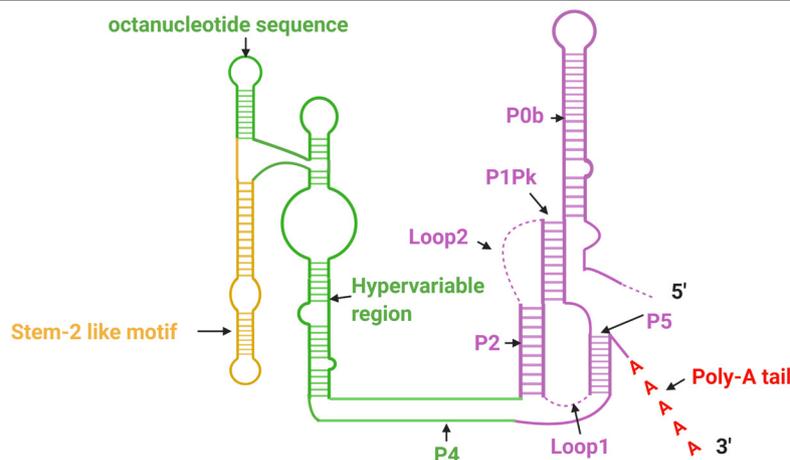


Figure 5. The secondary structure of 3' untranslated region (3'UTR) in SARS-CoV-2 RNA.

The hypervariable region consisting of octanucleotide sequence is shown in green, while the stem-2 like motif is shown in yellow. H-type pseudoknot with extended P0b stem is represented in purple. The poly(A) tail is indicated at the 3' terminus in red.

of an octanucleotide sequence 5'-GGAAGAGC-3', which deletion lowers pathogenicity of the virus in mice [49].

In other beta-coronaviruses, the s2m has been shown to form a bulged stem-loop (BSL), while the P1PK and hypervariable region have the same secondary structure as in SARS-CoV-2. Alignment based structural predictions suggest that the formation of P1PK requires the structural rearrangements to occur at the base of BSL, wherein the base pairing interactions can occur between the P1PK SL-2 and BSL 3' terminus. Further analysis has also revealed a short hairpin upstream to the P1PK SL 2, which partly overlaps with the P1PK loop 1 region and may compete with the base pairing interactions between P1PK and BSL [2].

Across all genera of human coronaviruses, the 3' UTR pseudoknot structure is phylogenetically conserved both in the location and structure, but only partly conserved in sequence. In particular, in beta-coronaviruses, i.e., HCoV-OC43, SARS-CoV, the pseudoknot and stem-loop struc-

tures are highly conserved, while alpha-coronaviruses, i.e., HCoV-229E and NL63 contain only conserved pseudoknot [30].

POST-TRANSCRIPTIONAL MODIFICATIONS OF CORONAVIRUS RNAs

Dynamic chemical modifications of viral RNAs, referred to as post-transcriptional modifications (PTM), play essential regulatory roles during the viral replication and pathogenesis [50]. These marks affect viral infectivity cycles in both the negative and positive manner. In ZIKV, DENV, and HIV 1, the modification processes promote the viral infectivity by facilitating viral replication [51], improving RNA stability [52], and upregulating the translation [53]. For other viruses, including HCV, measles virus, and respiratory syncytial virus, the epitranscriptomic marks negatively influence viral replication, terminate the synthesis of viral proteins, and prevent the production of progeny viral particles [54]. Post-transcriptional modifications have also

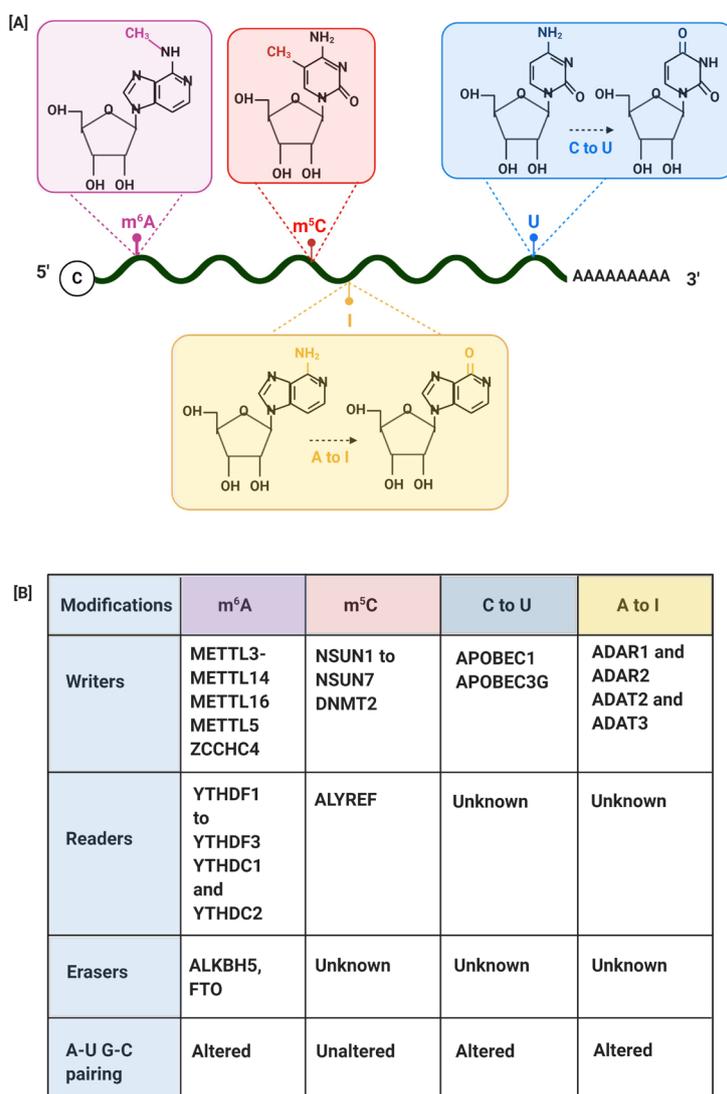


Figure 6. Post-transcriptional modifications identified in coronavirus RNA.

[A] Three types of post-transcriptional signatures have been identified within coronavirus transcriptome, and they included: m⁶A (purple box), m⁵C (red box), C-U editing (blue box), and A-I editing (yellow box). [B] The table lists specific writers, erasers, and readers proteins for each of the modifications, and the RNA structural effect of a given modification in the context of RNA pairing.

been proposed to facilitate the viral evasion of the cellular immune response. Viruses have been shown to expend the cellular epitranscriptomic machinery to mark their RNAs as “self” and prevent the recognition by RNA sensor melanoma differentiation-associated protein 5 [55]. Also, the modification of human metapneumovirus RNA allows it to escape RNA sensor, retinoic acid-inducible gene 1 (RIG-1) recognition, thus promoting viral replication [56].

There are over 140 chemical modifications to RNA, with N6-methyladenosine (m⁶A) and pseudouridine (Ψ) being the most prevalent in viral RNAs, as per recently performed mass-spectrometry analysis [57]. m⁶A has been shown to affect the RNA base pairing stability and exhibit varied hydrogen-bonding patterns to redefine higher-order RNA structure [58]. The methylation of cytosine (m⁵C) has little effect on the base pairing but is known to improve the major groove’s hydrophobicity and enhance the base stacking interactions [59]. The installation of Ψ affects RNA thermal stability, stacking interactions, and the base pairing between Ψ and any other nucleotides results in higher RNA structural stability [60]. The effects of A-to-I editing have been described as ‘unwinding activity’ on ds RNAs [61]. Also, the C-to-U editing causes changes in the pairing preference leading to the destabilization of ds regions [62].

The exploration of the epitranscriptomic landscapes of coronaviruses RNAs can provide valuable information for

identifying novel drug targets and optimizing the available therapeutics and mRNA-based vaccine development. Using nanopore DRS, it has been shown that SARS-CoV-2 genomic and sgRNAs can be decorated with 41 potential modifications, most of which localize to the AAGAA-like motif [20]. The overall frequency of current distortions corresponding to the modified sided has been noted at 20%, and it depends to some degree on the sgRNA species. In particular, the AAGAA-like motif is strongly modified within N sgRNA, while other types of distortions have been noted for ORF 3a, E, M sgRNAs. The authors have concluded that long viral transcripts, including genomic RNA, but also S, 3a, E, and M sgRNAs are modified more frequently than shorter sgRNAs, i.e., ORF 6, 7a-b, 8, and N. Also, it has been noted, that the modified RNAs have shorter poly(A) tails than unmodified transcripts, suggesting a link between the modifications and the functionality of the 3’ terminus. The authors speculated that because the poly(A) tail plays an essential function in RNA stability, the observed internal modification might be involved in controlling RNA turnover. It has been emphasized that the type of modification is yet to be identified. Furthermore, the comparison of DRS sequencing data for three SARS-CoV-2 patients isolates has validated the highest levels of modifications for S sgRNAs, followed with ORF 3a, E, M, ORF 6, 7a, 7b, ORF 8, and N. The authors noted that the RNA modification patterns are conserved in SARS-CoV-2 transcriptome and might be used for the

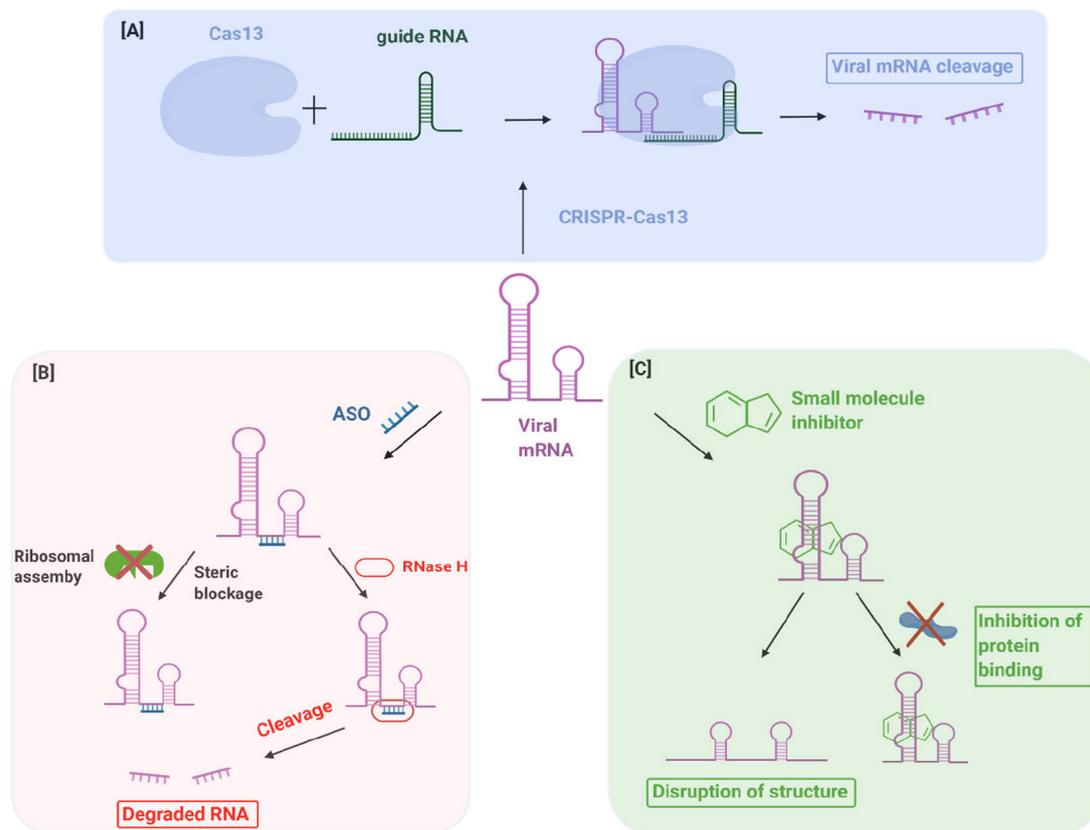


Figure 7. Therapeutic strategies for targeting viral RNA.

[A] The CRISPR-Cas13 system highlighted in blue box relies on the guide RNA (gRNA) which directs the Cas13 enzyme for cleavage of the target viral RNA. [B] Antisense oligonucleotide (ASO)-based strategy highlighted in pink box blocks the ribosomal assembly by binding to the target RNA, and also initiates the degradation of target RNA by activating the RNase pathway. [C] RNA-specific small molecule chemotype highlighted in green, binds to the RNA motif, and bring about conformational change resulting in the inhibition of viral activity.

identification of putative targets for the drug interventions. Another study indicated 42 positions with predicted m⁵C appearing consistently near the 3' termini of sgRNAs [63].

Certain RNA editing enzymes, e.g., apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like (APOBEC), and adenosine deaminase (ADAR), have been suggested to irreversibly re-code the primary sequences in SARS-CoV-2 genome [64]. The ADAR is known to act on ds RNA to deaminate adenines into inosines (A-to-I), while the APOBECs deaminate cytosines into uracils (C-to-U) on ss regions. The comparison of eight SARS-CoV-2 samples has led to the discovery of bias towards A>G transitions, which are frequently derived from A-to-I deamination. The second main group of chemical changes, namely C>T and G>A, has been proposed to derive from APOBEC-mediated deamination [64].

Nanopore sequencing has also been utilized for analyzing m⁵C patterns across the HCoV-229E transcriptome. The results indicated that both the TRS-L and the nested sgRNAs had a consistent pattern of methylation, suggesting that the process is sequence-specific and controlled by RNA structure. Unfortunately, the overall methylation pattern was similar to the negative control, i.e., unmethylated RNA calibration standard, and the false positive rate was calculated to be below 5% [65].

DRUGGABILITY OF CORONAVIRUSES RNA

Currently, there are no antiviral drugs or therapies that would show proven efficiency against coronaviruses infections in humans. This is, in part, due to the relatively high mutation rates observed for coronaviruses RNAs, which alter the core proteins, and in turn, lead to the rapid development of resistance against drugs and vaccines [6]. The discussed conserved *cis*-acting RNA motifs, critical for coronaviruses replication, transcription, packaging, and infectivity cycle, expanded the repertoire of potential antiviral targets. Here, the small molecules binding specifically to *cis*-acting RNA motif, and either disrupting its structure, altering its conformational flexibility or accessibility for intermolecular interactions, represent the unique therapeutic strategy. The chemical and biophysical tunability of small molecules, in addition to their excellent cellular permeability, has prompted the identification of ligand that binds to SARS-CoV FSE pseudoknot. *In silico* screening approach has identified 1,4-diazepane derivative, as a potent inhibitor of translational frameshifting in both *in vitro* and *in cellulo*-based assays. The optimal frameshifting rate is critical during the coronaviruses' infectivity cycle. Even a small difference in the percentage of frameshifting can have profound negative effects on viral propagation and infectivity. Further, surface plasmon resonance (SPR) target binding analysis showed that the binding of that ligand decreases the conformational plasticity of FSE fold [66]. Recent, homology-based sequence alignment studies of SARS-CoV-2 RNA aiming at the identification of potential therapeutic RNA targets, recognized 106 conserved structured regions that could be targeted with small molecules [25].

ASOs and small interfering RNA (siRNA) are also useful tools for RNA target validation and therapy. ASOs are ss oligonucleotides designed to target the complementary sequences and initiate degradation of target RNA by activating the endonuclease RNase H pathway [67]. Specific modifications to ASOs enhance their binding affinity pharmacokinetics, and tolerability profile, further expanding their therapeutic potential. For example, incorporating phosphonothioate, morpholino, or peptide nucleic acid into ASOs design improves their stability and cellular uptake [68]. Also, 2' ribose substitutions like 2' O-methyl (2' O-Me), 2' O-methoxyethyl (2'-MOE) and locked nucleic acid (LNA) enhance the target affinity and increase the resistance towards degradation by nucleases [67]. In comparison, siRNAs are a class of ds RNA molecules, 20-25 base pairs in length, which bind to the RNA induced silencing complex (RISC) and endonuclease Argonaute 2 (AGO2) complex, resulting in endonucleolytic cleavage of the target mRNA and gene silencing. Alike ASOs, various chemical modifications to siRNAs, including base modifications, phosphate backbone modifications, and sugar modifications, have been employed to improve their nuclease stability, binding affinity, and biodistribution of siRNA [69].

Previous work on SARS-CoV has indicated that N protein binding results in the unwinding of TRS-B structures that regulate the sgRNAs expression. Thus, a small molecule or ASO designed to bind and alter the TRS stability could hamper the expression of sgRNAs and act as an antiviral strategy [19]. Additionally, 59 conserved regions with low propensity to form stable structures have been predicted in SARS-CoV-2 RNA, and they all represent potential targets for oligonucleotide-based therapy [25]. The genome-wide structural study of SARS-CoV-2 has also revealed 261 accessible regions (21% of the genome), with 11 of them being located within ORF-N, which is found in every sgRNA [19]. Recent icSHAPE analysis provided insight into 469 regions, that are unlikely to form stable structures [26].

3D modeling of RNA structures can reveal distinct folds with conserved binding domains and druggable pockets for small molecules, thus providing an alternative approach for facilitating target discovery and development of antivirals. Secondary structure-restrained 3D modeling of the SARS-CoV-2 genome revealed putative druggable pockets within multi-way junctions and bulges. One of the promising and structurally well-defined motifs included the 3' UTR s2m, with a druggable pocket being located at the base of this structure [6].

Recent developments of the CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas13d-based system opened up the possibility of its application to identify and degrade the viral genome and virus-encoded RNA inside the living cell. A new CRISPR-Cas13d based strategy, prophylactic antiviral CRISPR in human cells (PAC-MAN), has been developed as a form of genetic intervention for the degradation of viral sequences by targeting highly conserved regions across SARS-CoV-2 and other human coronaviruses genomes. In the case of SARS-CoV-2, the highly conserved targets within the genome have been shown to

encode RDRP and N proteins, essential components for the coronavirus replication and function.

Recently, the bioinformatic analysis has been applied to identify sequences of CRISPR-associated RNAs (crRNAs) that can efficiently target coronavirus genomes. The study resulted in the identification of two crRNAs sequences that can target SARS-CoV-2, SARS, and MERS genomes, and six crRNAs that can target 91% of all sequenced coronaviruses genomes. The ability to use a small number of crRNAs that can target entire genomes highlights this method's unique power. Though PAC-MAN provided a proof of concept for targeting the conserved viral RNA sequences with repurposed RNA guided RNA endonuclease activity of Cas13, the development of potent and safe *in vivo* delivery methods is essential before it can be tested in clinical trials [7].

SUMMARY

Over the past decades, the emergence of many different coronaviruses that cause human diseases has occurred. These viruses will likely continue to emerge and evolve and cause human outbreaks due to their capacity to recombine, mutate, and infect multiple species and cell types. Gaining a comprehensive picture of the intricacies of the coronaviruses RNA genomes and encoded RNAs is thus necessary to provide a framework for developing novel and effective antiviral strategies.

Identifying the coronaviruses' *cis*-acting RNA motifs that have sufficient complexity, uniqueness, and can bind therapeutics with high affinity might be an exciting endeavor. However, the rules for determining the characteristics of functional inhibitors of RNA function are gradually being recognized. As per recent estimates [70], 431 RNA-targeting drug development programs, including mRNA vaccines, from Informa Pharma Intelligence's Biomedtracker, have been developed. Of these drug candidates, 63% are in the pre-Investigational New Drug (IND) stage, 32% are in early-stage clinical trials (phase I or II), 3% are in phase III, and five drugs are awaiting regulatory decisions. As such, whether an antiviral targeted to RNA is based on a small molecule or a large molecule such as antisense or RNAi, the perspective of establishing publicly available antivirals based on the viral RNA target now seems very close to reality. The next stage of targeting coronaviruses RNA must not only focus on the hunt for new antivirals and therapeutic strategies, but also on how to develop discovery platforms that give rise to lead compounds with high affinity, specificity, and bioavailability.

The expansion of additional tools and new refinements of existing methodologies, including advanced structural biology and modeling strategies for viral RNAs, can be leveraged to identify and rapidly characterize the subsection of *cis*-acting RNA motifs with high-quality ligand-binding domains and to rapidly evaluate the influence of ligand binding on the function of these sites. As we acquire more understanding of what constitutes a high-quality RNA target, RNA might prove to be no more problematic to drug than proteins. Given that targetable *cis*-acting RNA motifs appear to be abundant in viral RNA genomes and encoded

RNAs and that the rules dictating the RNA structure and dynamics are more defined, RNA might prove to be more broadly targetable than proteins.

REFERENCES

1. Khan S, Siddique R, Shereen MA, Ali A, Liu J, Bai Q, Bashir N, Xue M (2020) Emergence of a novel coronavirus, severe acute respiratory syndrome coronavirus 2: Biology and therapeutic options. *J Clin Microbiol* 58 :1-10
2. Madhugiri R, Fricke M, Marz M, Ziebuhr J (2014) RNA structure analysis of alphacoronavirus terminal genome regions. *Virus Res Elsevier BV* 194: 76-89
3. Liu DX, Liang JQ, Fung TS (2020) Human Coronavirus-229E, -OC43, -NL63, and -HKU1. *Ref Modul Life Sci* 43: 1-20
4. Sztuba-Solinska J, Chavez-Calvillo G, Cline SE (2019) Unveiling the druggable RNA targets and small molecule therapeutics. *Bioorganic Med Chem* 27: 2149-2165
5. Sztuba-Solinska J, Shenoy SR, Gareiss P, Krumpe LRH, Le Grice SFJ, O'Keefe BR, Schneekloth JS (2014) Identification of biologically active, HIV TAR RNA-binding small molecules using small molecule microarrays. *J Am Chem Soc* 136: 8402-8410
6. Manfredonia I, Nithin C, Ponce-Salvatierra A, Ghosh P, Wirecki TK, Marinus T, Snider EJ, Ogando NS, Hemert MJ Van, Bujnicki JM, Incarnato D (2020) Genome-wide mapping of therapeutically-relevant SARS-CoV-2 RNA structures. *Nucleic Acids Res*: 1-17
7. Abbott TR, Dhamdhare G, Liu Y, Lin X, Goudy L, Zeng L, Chemparathy A, Chmura S, Heaton NS, Debs R, Pande T, Endy D, La Russa MF, Lewis DB, Qi LS (2020) Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. *Cell Elsevier Inc* 181: 865-876
8. Singh J, Hanson J, Paliwal K, Zhou Y (2019) RNA secondary structure prediction using an ensemble of two-dimensional deep neural networks and transfer learning. *Nat Commun* 10: 5407
9. Leger A, Amaral P, Pandolfini L, Capitanchik C, Capraro F, Barbieri I, Migliori V, Luscombe N, Enright A, Tzelepis K, Ule J, Fitzgerald T, Birney E, Leonardi T, Kouzarides T (2019) RNA modifications detection by comparative Nanopore direct RNA sequencing. *bioRxiv*: 1-30
10. Huber R G, Lim XN, Ng WC, Sim AYL, Poh HX, Shen Y, Lim SY, Sundstrom KB, Sun X, Aw JG, Too HK, Boey PH, Wilm A, Cawla T, Choy MM, Jiang L, de Sessions PF, Loh XJ, Alonso S, Hibberd M, Nagarajan N, Ooi EE, Bond PJ, Sessions OM, Wan Y (2019) Structure mapping of dengue and Zika viruses reveals functional long-range interactions. *Nat Commun* 10: 1408
11. Mauger DM, Golden M, Yamane D, Williford S, Lemon SM, Martin DP, Weeks KM (2015) Functionally conserved architecture of hepatitis C virus RNA genomes. *Proc Natl Acad Sci* 112: 3692-3697
12. Liu P, Li L, Millership JJ, Kang H, Leibowitz JL, Giedroc DP (2007) A U-turn motif-containing stem-loop in the coronavirus 5' untranslated region plays a functional role in replication. *Rna* 13: 763-780
13. Plant EP, Pérez-Alvarado GC, Jacobs JL, Mukhopadhyay B, Hennig M, Dinman JD (2005) A three-stemmed mRNA pseudoknot in the SARS coronavirus frameshift signal. *PLoS Biol* 3: 1012-1023
14. Won C, Li L, Giedroc DP (2011) The solution structure of coronavirus stem-loop 2 (SL2) reveals a canonical CUYG tetraloop fold. *FEBS Lett* 585: 1049-1053
15. Madhugiri R, Karl N, Petersen D, Lamkiewicz K, Fricke M, Wend U, Scheuer R, Marz M, Ziebuhr J (2018) Structural and functional conservation of *cis*-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology* 517: 44-55.
16. Leamy KA, Assmann SM, Mathews DH, Bevilacqua PC (2016) Bridging the gap between *in vitro* and *in vivo* RNA folding. *Q Rev Biophys* 49:1-43
17. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM (2015) Selective 2'-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat Protoc* 10: 1643-1669

18. Sun L, Li P, Ju X, Rao J, Huang W, Zhang S, Xiong T, Xu K, Zhou X, Ren L, Ding Q, Wang J, Zhang QC (2020) In vivo structural characterization of the whole SARS-CoV-2 RNA genome identifies host cell target proteins vulnerable to re-purposed drugs. *bioRxiv*: 1-29
19. Lan TCT, Allan MF, Malsick L, Khandwala S, Nyeo S, Bathe M, Griffiths A, Rouskin S (2020) Structure of the full SARS-CoV-2 RNA genome in infected cells. *bioRxiv*: 1-28
20. Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H (2020) The Architecture of SARS-CoV-2 Transcriptome. *Cell Elsevier Inc* 181: 914-921
21. Ghut J, Aw A, Lim SW, Wang JX, Shen Y, Kaewsapsak P (2020) Direct RNA sequencing reveals structural differences between transcript isoforms. *bioRxiv*: 1-26.
22. Rother K, Rother M, Boniecki M, Puton T, Bujnicki JM (2011) RNA and protein 3D structure modeling: Similarities and differences. *J Mol Model* 17: 2325-2336
23. Rangan R, Watkins AM, Kladwang W, Das R (2020) De novo 3D models of SARS-CoV-2 RNA elements and small-molecule-binding RNAs to guide drug discovery. *bioRxiv*: 1-31
24. Andrews R, Peterson J, Haniff H, Chen J, Williams C, Grefe M, Disney M, Moss W (2020) An in silico map of the SARS-CoV-2 RNA Structure. *bioRxiv*: 1-14
25. Rangan R, Zheludev I, Hagey RJ, Pham EA, Wayment-Steele HK, Glenn JS, Das R (2020) RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: A first look. *RNA* 26:937-959
26. Sun L, Xu K, Huang W, Yang Y, Tang L, Xiong T, Zhang QC (2020) Predicting dynamic cellular protein-RNA interactions using deep learning and in vivo RNA structure. *bioRxiv*: 1-53
27. Schmidt N, Lareau C, Keshishian H, Melanson R, Zimmer M, Kirschner L, Ade J, Werner S, Caliskan N, Lander E, Vogel J, Carr SA, Bodem J, Munschauer M (2020) A direct RNA-protein interaction atlas of the SARS-CoV-2 RNA in infected human cells. *bioRxiv*: 1-33
28. Mchugh CA, Guttman M (2018) RAP-MS: A Method to Identify Proteins that Interact Directly with a Specific RNA Molecule in Cells. *Methods Mol Biol* 1649: 473-488.
29. Coleman CM, Frieman MB (2014) Coronaviruses: Important Emerging Human Pathogens. *J Virol* 88: 5209-5212
30. Yang D, Leibowitz JL (2015) The structure and functions of coronavirus genomic 3' and 5' ends. *Virus Res* 206: 120-133
31. Sola I, Almazán F, Zúñiga S, Enjuanes L (2015) Continuous and Discontinuous RNA Synthesis in Coronaviruses. *Annu Rev Virol* 2: 265-288
32. Sola I, Mateos-Gomez PA, Almazan F, Zúñiga S, Enjuanes L (2011) RNA-RNA and RNA-protein interactions in coronavirus replication and transcription. *RNA Biol* 8: 237-248
33. Mishra A, Pandey AK, Gupta P, Pradhan P, Dhamija S, Gomes J, Kundu B, Vivekanandan P, Menon MB (2020) Mutation landscape of SARS-CoV-2 reveals three mutually exclusive clusters of leading and trailing single nucleotide substitutions. *bioRxiv*: 1-19
34. Wu HY, Brian DA (2010) Subgenomic messenger RNA amplification in coronaviruses. *Proc Natl Acad Sci USA* 107: 12257-12262
35. Sanders W, Fritch EJ, Madden EA, Graham RL, Vincent HA (2020) Comparative analysis of coronavirus genomic RNA structure reveals conservation in SARS-like coronaviruses. *bioRxiv*: 1-14
36. Wakida H, Kawata K, Yamaji Y, Hattori E, Tsuchiya T, Wada Y, Ozaki H, Akimitsu N (2020) Stability of RNA sequences derived from the coronavirus genome in human cells. *Biochem Biophys Res Commun* 527: 993-999
37. Incarnato D, Anselmi F, Morandi E, Neri F, Maldotti M, Rapelli S, Parlato C, Basile G, Oliviero S (2017) High-throughput single-base resolution mapping of RNA 2'-O-methylated residues. *Nucleic Acids Res* 45:1433-1441
38. Thapar R, Denmon AP, Nikonowicz EP (2014) Recognition modes of RNA tetraloops and tetraloop-like motifs by RNA-binding proteins. *Wiley Interdiscip Rev RNA* 5: 49-67
39. Siegfried NA, Busan S, Rice GM, Nelson JAE, Weeks KM (2014) RNA motif discovery by SHAPE and mutational profiling (SHAPE-MaP). *Nat Methods* 11: 959-965
40. Madhugiri R, Karl N, Petersen D, Lamkiewicz K, Fricke M, Wend U, Scheuer R, Marz M, Ziebuhr J (2018) Structural and functional conservation of cis-acting RNA elements in coronavirus 5'-terminal genome regions. *Virology* 517: 44-55
41. Kelly JA, Olson AN, Neupane K, Munshi S, San Emeterio J, Pollack L, Woodside MT, Dinman JD (2020) Structural and functional conservation of the programmed-1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *J Biol Chem* 295: 10741-10748
42. Cho C, Lin S, Chou M, Hsu H, Chang K (2020) Regulation of programmed ribosomal frameshifting by co-translational refolding of rna hairpins. *PLoS one* 8: 1-11
43. Herold J, Siddell SG (1993) An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res* 21: 5838-5842
44. Yang D, Liu P, Giedroc DP, Leibowitz J (2011) Mouse hepatitis virus stem-loop 4 functions as a spacer element required to drive subgenomic RNA synthesis. *J Virol* 85: 9199-9209
45. Mustoe AM, Busan S, Rice GM, Hajdin CE, Peterson BK, Ruda VM, Kubica N, Nutiu R, Baryza JL, Weeks KM (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173: 181-195
46. Nakagawa K, Lokugamage KG, Makino S (2016) Viral and cellular mRNA translation in coronavirus-infected cells. *Adv Virus Res* 96: 165-192
47. Goebel SJ, Hsue B, Dombrowski TF, Masters PS (2004) Characterization of the RNA components of a putative molecular switch in the 3' untranslated region of the murine coronavirus genome. *J Virol* 78: 669-682
48. Robertson MP, Igel H, Baertsch R, Haussler D, Ares M, Scott WG (2005) The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Biol* 3: 1-9
49. Goebel SJ, Miller TB, Bennett CJ, Bernard KA, Masters PS (2007) A hypervariable region within the 3' cis-acting element of the murine coronavirus genome is nonessential for RNA synthesis but affects pathogenesis. *J Virol* 81: 1274-1287
50. Gonzales-van Horn SR, Sarnow P (2017) Making the mark: the role of adenosine modifications in the life cycle of RNA viruses. *Cell Host Microbe* 21: 661-669
51. Lichinchi G, Zhao BS, Wu Y, Lu Z, Qin Y, He C, Rana TM (2016) Dynamics of human and viral RNA methylation during Zika virus infection. *Cell Host Microbe* 20: 666-673
52. Kennedy EM, Courtney DG, Tsai K, Cullen BR (2017) Viral epitranscriptomics. *J Virol* 91: 1-10
53. Courtney DG, Chalem A, Bogerd HP, Law BA, Kennedy EM, Holley CL, Cullen BR (2019) Extensive epitranscriptomic methylation of A and C residues on murine leukemia virus transcripts enhances viral gene expression. *MBio* 10: 1-12
54. Gokhale NS, McIntyre ABR, McFadden MJ, Roder AE, Kennedy EM, Gandara JA, Hopcraft SE, Quicke KM, Vazquez C, Willer J, Ilkayeva OR, Law BA, Holley CL, Garcia-Blanco MA, Evans MJ, Suthar MS, Bradrick SS, Mason CE, Horner SM (2016) N6-Methyladenosine in flaviviridae viral RNA genomes regulates infection. *Cell Host Microbe* 20: 654-665
55. Züst R, Cervantes-Barragan L, Habjan M, Maier R, Neuman BW, Ziebuhr J, Szretter KJ, Baker SC, Barchet W, Diamond MS, Siddell SG, Ludewig B, Thiel V (2011) Ribose 2'-O-methylation provides a molecular signature for the distinction of self and non-self mRNA dependent on the RNA sensor Mda5. *Nat Immunol* 12: 137-143
56. Tsai K, Cullen BR (2020) Epigenetic and epitranscriptomic regulation of viral replication. *Nat Rev Microbiol* Springer US 18: 559-570
57. McIntyre W, Netzband R, Bonenfant G, Biegel JM, Miller C, Fuchs G, Henderson E, Arra M, Canki M, Fabris D, Pager CT (2018) Positive-sense RNA viruses reveal the complexity and dynamics of the cellular and viral epitranscriptomes during infection. *Nucleic Acids Res* 46: 5776-5791

58. Liu B, Merriman DK, Choi SH, Schumacher MA, Plangger R, Kreutz C, Horner SM, Meyer KD, Al-Hashimi HM (2018) A potentially abundant junctional RNA motif stabilized by m6A and Mg²⁺. *Nat Commun* 9: 2761
59. Wang S, Kool ET (1995) Origins of the large differences in stability of DNA and RNA helices: c-5 methyl and 2'-hydroxyl effects. *Biochemistry* 34: 4125–4132
60. Hudson GA, Bloomingdale RJ, Znosko BM (2013) Thermodynamic contribution and nearest-neighbor parameters of pseudouridine-adenosine base pairs in oligoribonucleotides. *Rna* 19: 1474–1482
61. Bass BL, Weintraub H (1987) A developmentally regulated activity that unwinds RNA duplexes. *Cell* 48: 607–613
62. Harcourt EM, Kietrys AM, Kool ET (2017) Chemical and structural effects of base modifications in messenger RNA. *Nature* 541: 339–346
63. Taiaroa G, Rawlinson D, Featherstone L, Pitt M, Caly L (2020) Direct RNA sequencing and early evolution of SARS-CoV-2. *bioRxiv*: 1-17
64. Giorgio SDi, Martignano F, Torcia MG, Mattiuz G, Conticello SG (2020) Evidence for host-dependent RNA editing in the transcriptome of SARS-CoV-2. *Sci Adv* 6: 5813
65. Viehweger A, Krautwurst S, Lamkiewicz K, Madhugiri R, Ziebuhr J, Hölzer M, Marz M (2019) Direct RNA nanopore sequencing of full-length coronavirus genomes provides novel insights into structural variants and enables modification analysis. *Genome Res* 29: 1545–1554
66. Hermann T (2016) Small molecules targeting viral RNA. *Wiley Interdiscip Rev RNA* 7: 726–743
67. Wacheck V, Zangemeister-Wittke U (2006) Antisense molecules for targeted cancer therapy. *Crit Rev Oncol Hematol* 59: 65–73
68. Schoch KM, Miller TM (2017) Antisense Oligonucleotides: Translation from Mouse Models to Human Neurodegenerative Diseases. *Neuron* 94: 1056–1070
69. Chery J (2016) RNA therapeutics: RNAi and antisense mechanisms and clinical applications. *Postdoc J* 4: 35–50
70. Wang F, Zuroske T, Watts JK (2020) RNA therapeutics on the rise. *Nat Rev Drug Discov* 19: 441–442

Zgłębiając zawiłości struktur, funkcji i lekowności cis motywów RNA u koronawirusów

Joanna Sztuba-Solinska* and Devadatta Gosavi

Department of Biological Sciences, Auburn University, 120 W. Samford Ave, Rouse Life Sciences Building, Auburn, AL 36849, United States

*autor korespondujący: tel: 334-844-1650; email: jzs0165@auburn.edu

STRESZCZENIE

Rodzina Koronawirusów obejmuje otoczkowe, jednoniciowe (ss+) wirusy RNA powodujące łagodne lub ciężkie infekcje dróg oddechowych i jelit u ludzi. Wirus SARS-CoV-2, który zaatakował ludzkość w ubiegłym roku, jest przyczyną zespołu ostrej niewydolności oddechowej (COVID-19) i doprowadził do śmierci miliony ludzi na całym świecie. Zatrzymanie pochodzącego wirusa jest możliwe tylko dzięki odkryciu i wdrożeniu skutecznej terapii przeciwwirusowej. Zdolność koronawirusów do szybkiej ewolucji, adaptacji i przekraczania barier międzygatunkowych sprawia, że opracowanie takich standardów terapeutycznych stanowi pilną potrzebę. Postępy w dziedzinie biologii RNA, a zwłaszcza wykorzystanie narzędzi do głębokiego sekwencjonowania dostarczyły ostatnio wszechstronnych informacji na temat struktur drugo- i trzeciorzędowych przyjmowanych przez RNA koronawirusów. Uzyskanie kompleksowego obrazu materiału genetycznego koronawirusów daje nadzieję na opracowanie nowych i skutecznych strategii przeciwwirusowych. W tym manuskrypcie przedstawiamy szeroki przegląd motywów i domen regulatorowych RNA (*cis-acting*) u ludzkich koronawirusów. Omawiamy ich funkcjonalność, strukturę, modyfikacje potranskrypcyjne i oceniamy potencjał ich wykorzystania jako celów terapeutycznych. Koncentrujemy się na najnowszych pracach, które dostarczyły znaczącego wglądu w konformację strukturalną RNA u SARS-CoV-2, a także odwołujemy się do poprzednich kluczowych badań dotyczących struktury i funkcji RNA innych ludzkich koronawirusów. Opisujemy również podstawy metodologii stosowanej do globalnej oceny wirusowego RNA począwszy od spektroskopii magnetycznego rezonansu jądrowego (NMR), poprzez chemiczne mapowanie struktur metodami SHAPE-MaP, icSHAPE, DMS-MaPseq, bezpośrednio sekwencjonowanie RNA przy użyciu systemu MinION, aż po modelowanie *in silico* z użyciem algorytmów FARNa, FARFAR, ScanFold, RNAz i Contrafold. Celem poniższej publikacji jest zebranie cennego materiału źródłowego, który ma stanowić inspirację do prowadzenia dalszych badań nad strukturą i funkcją RNA koronawirusów, a także nad poszukiwaniem metod leczenia COVID-19.