# Introduction

The emergence of coronavirus disease 2019 (COVID-19) as a pandemic in 2020 spurred many investigations on functional RNA structures in coronaviruses, particularly SARS coronavirus 2 (SARS-CoV-2) [1, 2, 3, 4, 5, 6, 7, 8, 9, 10]. Among the more unexpected findings was an RNA:RNA interaction between the frameshifting stimulation element (FSE) and another sequence up to 1,475 nt downstream, which the authors named the FSE-arch [3]. The FSE-arch was detected in infected cells using COMRADES [11] and proposed to comprise three nested long-range RNA:RNA interactions (Figure ??a): an outer 38 bp bulged stem spanning coordinates 13,370-14,842 (which encompasses the FSE); a middle 18 bp bulged stem spanning coordinates 13,533-14,673; and an inner 14 bp bulged stem spanning coordinates 13,580-14,552 [3]. We had discovered that the FSE folds into at least two alternative structures in infected cells, in roughly equal proportions, and that the predicted structure for one of them resembles the FSE-arch [10]. Because computational RNA structure prediction – even guided by chemical probing data – is unreliable for long RNA sequences especially [12], we sought stronger, hypothesis-driven evidence for the existence of the FSE-arch.

Chemical probing followed by mutational profiling is a common strategy for inferring secondary structures of RNA molecules [13, 14].

Here, we present a method to probe RNA–RNA interactions spanning hundreds to thousands of nucleotides, "Structure Ensemble Ablation by Reverse Complement Hybridization with Mutational Profiling" (SEARCH-MaP). To compute, compare, and deconvolute data from mutational profiling experiments (including SEARCH-MaP, DMS-MaPseq, and SHAPE-MaP), we introduce the software "Structure Ensemble Inference by Sequencing, Mutation Identification, and Clustering of RNA" (SEISMIC-RNA).

Results

Strategy of SEARCH-MaP and SEISMIC-RNA
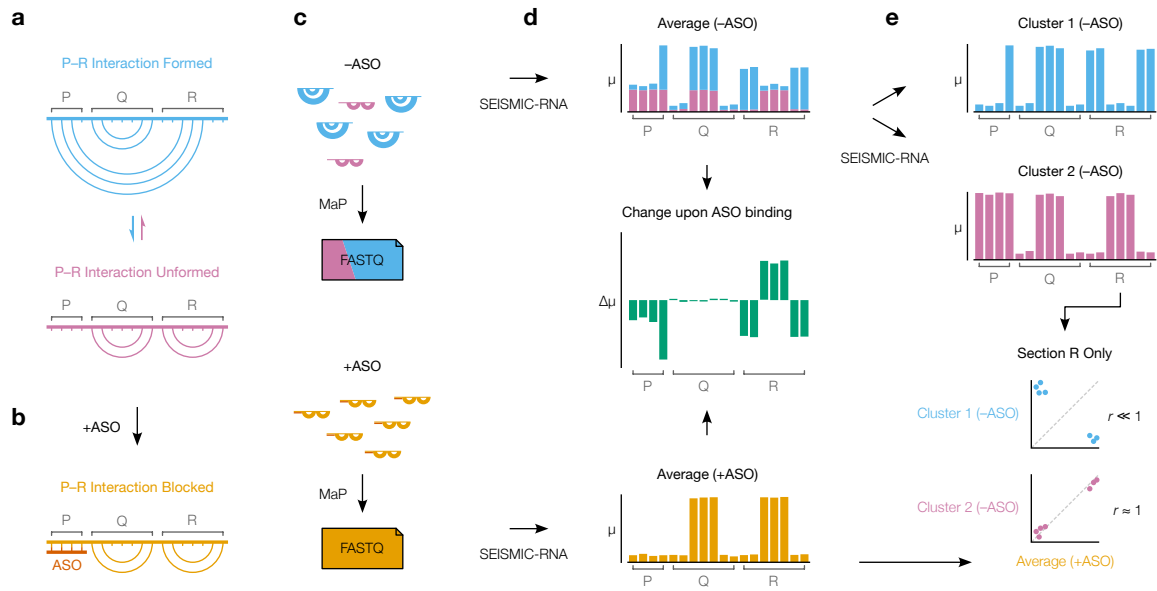
Figure 1: **The strategy of SEARCH-MaP and SEISMIC-RNA. (a)** This toy RNA is partitioned into three sections (P, Q, and R) whose molecules exist in two structural states: one in which an interaction between P and R forms (blue) and one in which it does not (purple). **(b)** Hybridizing an ASO (red) to P blocks it from interacting with R and forces all RNA molecules into the state where the P–R interaction is unformed. **(c)** A SEARCH-MaP experiment entails separate chemical probing and mutational profiling (MaP) with (+ASO) and without (–ASO) the ASO, followed by sequencing to generate FASTQ files. The RNA molecules and FASTQ files use the same color scheme as in (a) and are illustrated/colored in proportion to their abundances in the ensemble. **(d)** Ensemble average mutational profiles with (+ASO) and without (–ASO) the ASO, computed with SEISMIC-RNA. The *x*-axis is the position in the RNA sequence; the *y*-axis is the fraction of mutations (μ) at the position. Each bar in the –ASO profile is drawn in two colors merely to illustrate how much each structural state contributes to each position; in a real experiment, states cannot be distinguished before clustering. The change upon ASO binding (green) indicates the difference in the fraction of mutations ($\Delta\mu$) between the +ASO and –ASO conditions. **(e)** Mutational profiles of two clusters (top) obtained by clustering the –ASO ensemble in (d) using SEISMIC-RNA, and the scatter plot of the mutation rates of bases in R (bottom) between the +ASO ensemble average (*x*-axis) and each cluster (*y*-axis). The expected correlation (r) is shown beside each scatter plot.

We illustrate SEARCH-MaP with an RNA comprising three sections (P, Q, and R) that folds into an ensemble of two structural states: one in which a base-pairing interaction between P and R forms, another in which it does not (Figure 1a). Searching for sections that interact with P begins with hybridizing an antisense oligonucleotide (ASO) to P,

which blocks P from base pairing with any other section, ablating the state in which the P–R interaction forms (Figure 1b). The RNA is chemically probed separately with (+ASO) and without (–ASO) the ASO, followed by mutational profiling and sequencing, e.g. using DMS-MaPseq [13] (Figure 1c).

SEISMIC-RNA can detect RNA–RNA interactions by comparing the +ASO and –ASO mutational profiles. Theoretically, each structural state has its own mutational profile [15], but the mutational profile of a single state is not directly observable because all states are physically mixed during the experiment (Figure 1c, top). Instead, the directly observable mutational profile is the "ensemble average" – the average of the states' (unobserved) mutational profiles, weighted by the states' (unobserved) proportions (Figure 1d, top). Because the structures – and therefore mutational profiles – of R differ between the interaction-formed and -unformed states, the ensemble averages of R also differ between the +ASO and –ASO conditions (Figure 1d, middle). However, this is not the case for element Q, which has the same secondary structure in both states (Figure 1d, middle). Therefore, one can deduce that P interacts with R – but not with Q – because hybridizing an ASO to P alters the mutational profile of R but not of Q.

After identifying RNA–RNA interactions, SEISMIC-RNA can also determine the mutational profiles of the states where the P–R interaction is formed and unformed – even if their secondary structures are unknown. Inferring mutational profiles for the interaction-formed and -unformed states requires clustering the –ASO ensemble into two clusters of RNA molecules (Figure 1e, top). Each cluster has its own mutational profile and corresponds to one structural state, but which cluster corresponds to the interaction-formed (or -unformed) state is not yet known. The interaction-unformed state has a mutational profile similar to that of the +ASO ensemble average, since the ASO blocks the interaction and forces the RNA into the interaction-unformed state. Therefore, a cluster that correlates well ($r \approx 1$) with the +ASO ensemble average (here, Cluster 2) corresponds to the interaction-unformed state; while a cluster that correlates weakly ($r \ll 1$) corresponds to the interaction-formed state (Figure 1e, bottom).

(I hope) SEARCH-MaP detects long-range base-pairing in ribosomal RNA

We first validated SEARCH-MaP using 16S and 23S ribosomal RNA (rRNA) from *E. coli*. For each rRNA, we selected two RNA–RNA interactions spanning ¿ [HOW MANY] nt that had been detected in a cell-free system [16]. For each interaction, we hypothesized that binding an ASO to either side would break the interaction and perturb the structure of the other side (distant from the ASO binding site) and designed two ASOs, one targeting each side. As a negative control, we also designed one ASO targeting a stem loop in each rRNA, which we hypothesized would perturb only the structure near the ASO binding site.

We folded the 16S and 23S rRNAs with each ASO, performed DMS-MaPseq over the entire transcripts, and compared ensemble average mutational profiles with and without ASOs using SEISMIC-RNA. [DESCRIBE THE RESULTS]
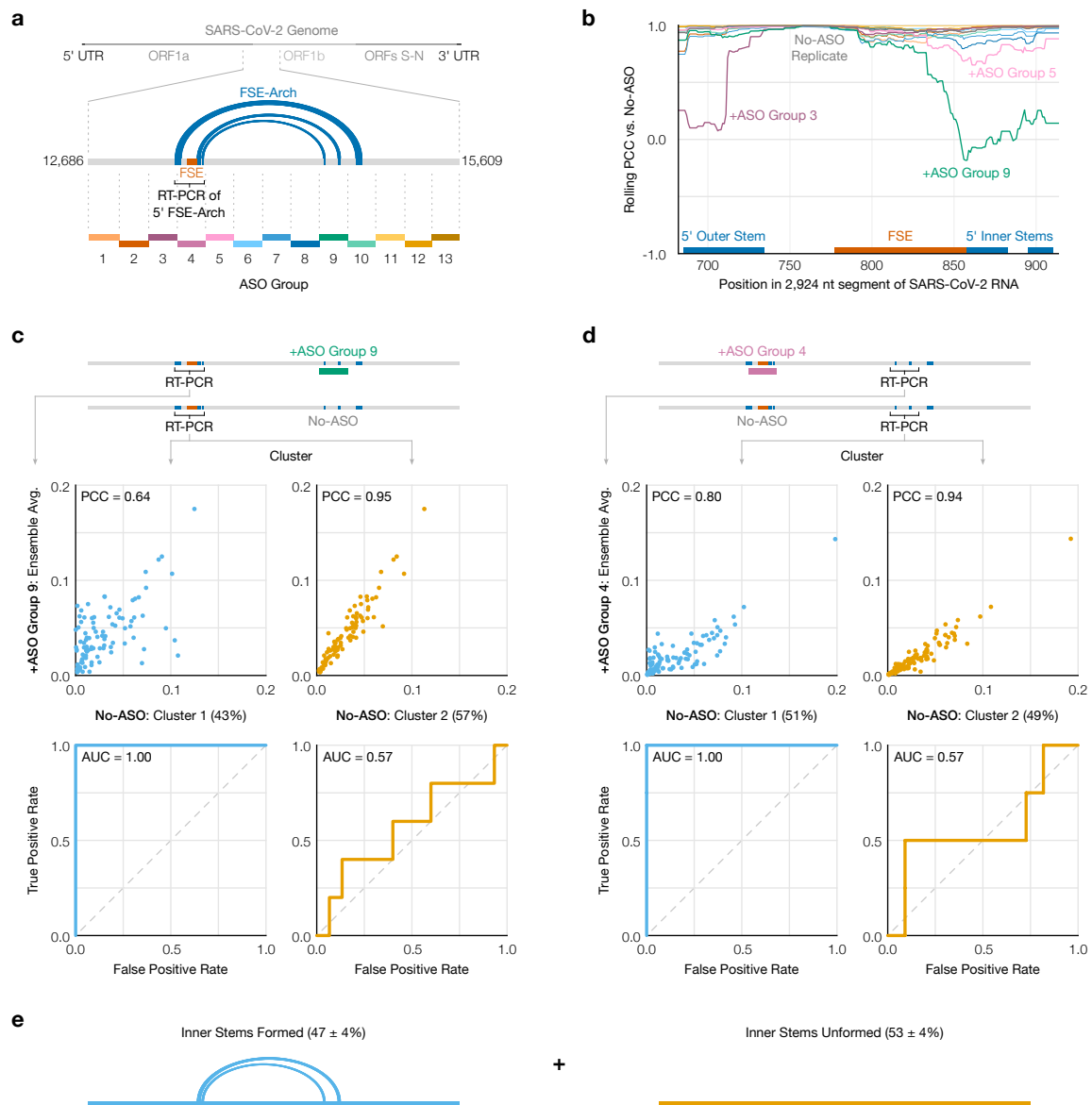
Figure 2:

SEARCH-MaP detects, separates, and quantifies a long-range RNA–RNA interaction in SARS-CoV-2

Aside from ribosomes, many of the best-characterized functional long-range RNA–RNA interactions occur in the genomes of RNA viruses [17]. Coronaviruses regulate translation of their first open reading frame (ORF1) using programmed ribosomal frameshifting [18]. In the middle of ORF1, a switch called a frameshift stimulation element (FSE) makes a fraction of ribosomes slip backwards into the -1 reading frame. Ribosomes that maintain reading frame terminate at a stop codon shortly after the FSE, while those that frameshift bypass that stop codon and reach the end of ORF1. Why coronaviruses need a frameshifting mechanism remains an open question [19], yet all have FSEs [18].

Every coronaviral FSE contains a "slippery site" (UUUAAAC) and a structure characterized as a pseudoknot in multiple species [20, 21, 22]. Indeed, the isolated core of the SARS coronavirus 2 (SARS-CoV-2) FSE was shown to fold into a pseudoknot with three stems [23, 24]. However, we discovered that when FSE is in its natural place in

the SARS-CoV-2 genome, pseudoknot stem 1 is disassembled while an alternative stem 1 folds [10]. A 283 nt segment of the RNA genome – containing both the FSE and alternative stem 1 – failed to fully mimic the DMS reactivities of the full virus (PCC = 0.75). A 2,924 nt segment came closer (PCC = 0.93), suggesting that – only in the context of this longer sequence – the FSE adopts yet another structure, presumably a long-range interaction [10].

We used SEARCH-MaP to find the long-range interaction involving the FSE. We hypothesized it would turn out to be the structure another group had discovered and named the "FSE-arch" [3]. If so, the structure of the FSE (at the 5' side of the FSE-arch) would be perturbed by – and only by – ASOs targeting either side of the putative FSE-arch. To investigate, we added (separately) thirteen groups of DNA ASOs to the 2,924 nt segment (Figure 3a). Each group contained four or five ASOs targeting a contiguous 213-244 nt section of the RNA. After adding each group of ASOs, we performed DMS-MaPseq with two pairs of RT-PCR primers: flanking the ASO target site (to confirm binding) and flanking the 5' FSE-arch (to detect structural changes). We obtained data for every ASO group except 13. Every ASO group bound properly, evidenced by suppression of DMS reactivities over its target site (SFIG).

Figure 3: **Search for a long-range RNA–RNA interaction with the SARS-CoV-2 FSE.**
(a) The 2,924 nt segment of the SARS-CoV-2 genome containing the frameshift stimula-
tion element (FSE) and putative FSE-arch [3]. The target of each ASO group is indicated
by dotted lines; lengths are to scale. **(b)** Rolling (window = 45 nt) Pearson correlation
coefficient (PCC) of DMS reactivities over the 5' FSE-arch after adding each ASO group
versus a no-ASO control. Each curve represents one ASO group, colored as in (a); groups
4 and 13 are not shown. Locations of the FSE and the outer and inner stems of the 5'
FSE-arch are also indicated. **(c)** (Top) Scatter plots of DMS reactivities over the 5' FSE-
arch comparing each cluster of the no-ASO sample to the sample with ASO group 9;
each point is one position in the 5' FSE-arch. (Bottom) Receiver operating characteristic
(ROC) curves comparing each cluster of the no-ASO sample to the two inner stems of
the FSE-arch. **(d)** Like (c) but over the 3' FSE-arch, and comparing to the sample with
ASO group 4. One highly reactive outlier was ignored when calculating PCC. **(e)** Model
of the inner two stems in the ensemble of structures formed by the 2,924 nt segment.

To quantify structural changes over the 5' FSE-arch, we calculated the rolling Pearson

correlation coefficient (PCC) of the DMS reactivities between each sample and a no-ASO

control (Figure 3b). A no-ASO replicate had a rolling PCC consistently between 0.93 and 1.00 (mean = 0.97), confirming the DMS reactivities were reproducible. ASO group 9 – targeting both 3' inner stems of the FSE-arch – caused the rolling PCC to dip below 0.5 near the 5' inner stems, exactly as expected if the inner stems of the FSE-arch existed. The only other ASO groups with substantial effects were 3, 4, and 5, which overlapped or abutted the FSE and presumably perturbed short-range base pairs; the outer stem of the FSE-arch (targeted by ASO group 10) did not apparently form. These results suggest both inner stems of the FSE-arch exist and are the predominant long-range interactions involving the immediate vicinity of the FSE.

We next sought to determine in what fraction of molecules the two inner stems of the FSE-arch form. Using SEISMIC-RNA, we clustered reads from the 5' side of the FSE-arch for the no-ASO control and found two distinct clusters, with a 43%/57% split. To determine if these clusters corresponded to the two inner stems formed and unformed, we compared their DMS reactivities to those after adding ASO group 9, which blocks the two inner stems (Figure 3c, top). Cluster 2 had similar DMS reactivities (PCC = 0.95), indicating it corresponds to the stems unformed. Meanwhile, the DMS reactivities of cluster 1 differed (PCC = 0.64), suggesting it corresponds to the stems formed.

To further verify this correspondence, we leveraged the preexisting model of the FSE-arch [3]. If cluster 1 did correspond to the two inner stems formed, then we would expect its DMS reactivities to agree well with their structures (i.e. paired and unpaired bases should have low and high reactivities, respectively) and those of cluster 2 to agree much less. We quantified this agreement using receiver operating characteristic (ROC) curves (Figure 3c, bottom). The area under the curve (AUC) for cluster 1 was 1.00, indicating perfect agreement with the two inner stems of the FSE-arch; while that of cluster 2 was 0.57, close to null (0.50). This result further supports that cluster 1 (43%) corresponds to the two inner stems formed, and cluster 2 (57%) to these stems unformed.

If the RNA really exists as an ensemble of the two inner stems formed and unformed, then we would also expect the 3' side of the FSE-arch to cluster into formed and un-formed states. To investigate, we performed RT-PCR with primers flanking the 3' side of the inner two stems – both without ASOs and with ASO group 4 (targeting the 5' side of the FSE-arch). We clustered the no-ASO control and found – similar to the previous result – that cluster 1 (51%) matched the structure of the inner two stems (AUC = 1.00),

while cluster 2 (49%) agreed with the DMS reactivities (PCC = 0.94) after blocking the FSE-arch with ASO group 4 (Figure 3d). We concluded that the RNA exists as an ensemble of structures in which the two inner stems of the FSE-arch form in 47% ± 4% of molecules (Figure 3e).

The long-range interaction competes with the frameshift pseudoknot in SARS-CoV-2

Having clustered out the DMS reactivities of the interaction-formed state on both sides of the FSE-arch (cluster 1 in Figure 3c and d), we used them as DMS constraints [25] in RNAstructure [26] to fold a 1,799 nt segment centered on the long-range interaction. This refined model (Figure 4) contained not only the two inner stems of the FSE-arch – which we hereafter call long stems 1 (LS1) and 2 (LS2) – but also two stems (LS3 and LS4) that were not part of the original FSE-arch model [3]. The structure also contained the alternative stem 1 (AS1) that we had previously discovered [10]. To our surprise, LS2b, LS3, and LS4 of the new model overlapped all three stems of the pseudoknot (PS1, PS2, and PS3) that is generally thought to stimulate frameshifting [27, 23, 24]. Thus, these long stems – if they exist – and the pseudoknot would be mutually exclusive.
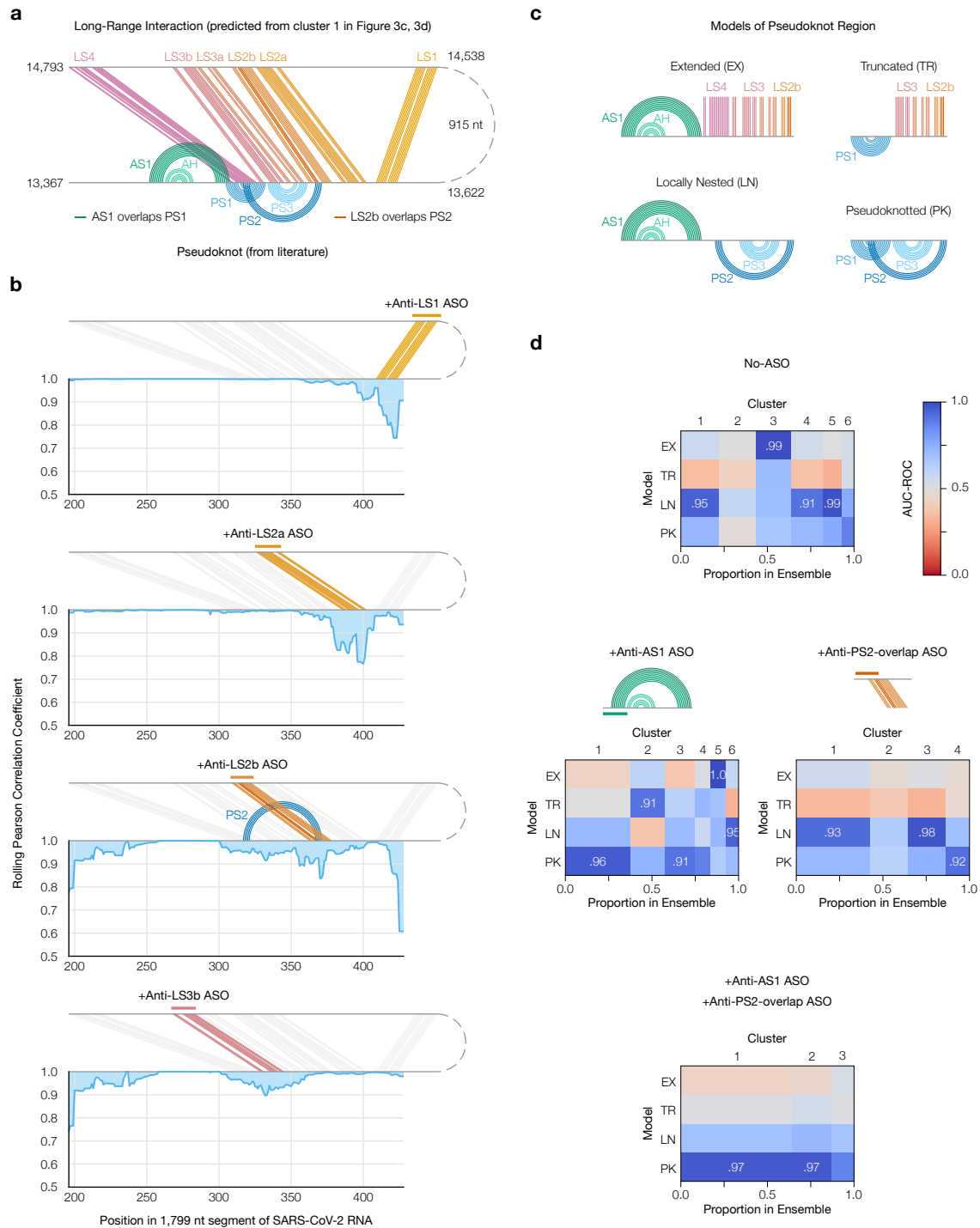
Figure 4: **Refinement of the long-range interaction and competition with the frameshift pseudoknot. (a)** Refined model of the long-range interaction (predicted from cluster 1 in Figure 3c and d) including alternative stem 1 (AS1) [10]; the attenuator hairpin (AH) [28]; and long stems LS1, LS2a/b, LS3a/b, and LS4. Locations of pseudoknot stems PS1, PS2, and PS3 are also shown; as are the base pairs they overlap in AS1 and LS2b. **(b)** Rolling (window = 21 nt) Pearson correlation coefficient of DMS reactivities after adding each ASO versus a no-ASO control; base pairs targeted by each ASO are colored. **(c)** Models of possible structures for the FSE, by combining mutually compatible stems from (a). **(d)** Heatmaps comparing models in (c) to clusters of DMS reactivities over positions 305-371 via the area under the receiver operating characteristic curve (AUC-ROC). AUC-ROCs of at least 0.90 are annotated. Cluster widths indicate proportions in the ensemble.

To verify the newly predicted long stems, we performed SEARCH-MaP on the 1,799 nt segment using 15-20 nt LNA/DNA mixmer ASOs to reach single-stem precision (Figure 4b). Each ASO targeted one stem in the downstream portion of the interaction, and we measured the change in DMS reactivities of the FSE. ASOs targeting the 3' sides of LS1 and LS2a perturbed the DMS reactivities in exactly the expected locations on the 5' sides. Binding an ASO to the 3' side of LS2b caused a larger perturbation with more off-target effects, likely because this stem overlaps with pseudoknot stem 2 (PS2). Blocking LS3b also resulted in a main effect around the intended location, with one off-target effect upstream, suggesting that there may be another RNA–RNA interaction with the pseudoknot and this upstream region. Therefore, stems LS1, LS2a/b, and LS3b do exist – at least in a portion of the ensemble.

We then sought to determine whether the long-range stems compete with the pseudoknot. If so, blocking them with ASOs would increase the proportion of the pseudoknot in the ensemble. To test this hypothesis, we first generated four possible models of the FSE structure by combining mutually compatible stems from the refined model (Figure 4c). Then, we clustered the 1,799 nt segment without ASOs up to 6 clusters (the maximum number reproducible between replicates) and compared each cluster to each structure model using the area under the receiver operating characteristic curve (AUC-ROC) over the positions spanned by the pseudoknot, 305-371 (Figure 4d, top). We considered a cluster and model to be consistent if the AUC-ROC was at least 0.90. The locally nested model (AS1 plus PS2 and PS3) was consistent with three clusters totaling 52% of the ensemble, while the extended model (AS1 plus all long-range stems) was consistent with one cluster (20%). No clusters were fully consistent with the pseudoknotted model, though the least-abundant cluster (7%) came close with an AUC-ROC of 0.88. The remaining cluster (21%) was not consistent with any model, suggesting that the ensemble contains structures beyond those in Figure 4c.

Adding an ASO targeting the 5' side of AS1 reduced the proportion of AS1-containing states (extended and locally nested) from 72% to 16% (Figure 4d, left). In their absence emerged clusters consistent with the pseudoknotted and truncated models, representing 56% and 20% of the ensemble, respectively. Meanwhile, adding an ASO that blocked the part of LS2b that overlaps PS2 eliminated the extended state (which includes LS2b) and produced one cluster (13%) consistent with the pseudoknotted model (Figure 4d,

right). Adding both ASOs simultaneously ablated the extended, truncated, and locally nested states, collapsing the ensemble into three clusters of which two (87%) were highly consistent with the pseudoknotted model (Figure 4d, bottom). Since blocking the PS2-overlapping portion of LS2b increased the proportion of clusters consistent (or nearly so) with the pseudoknotted model – both alone and combined with the anti-AS1 ASO – we conclude that the long-range interaction does outcompete the pseudoknot.

Frameshift stimulating elements of multiple coronaviruses participate in long-range RNA–RNA interactions

To search for long-range interactions involving the FSE in other coronaviruses, we performed SEARCH-MaP with FSE-targeted ASOs on 1,799 nt segments from eight coronaviral genomes.

Computational and experimental screening identifies eight coronaviruses with potential long-range interactions

As of December 2021, the NCBI Reference Sequence Database [29] contained 62 complete genomes of coronaviruses. To focus on those likely to have long-range interactions involving the FSE, we predicted the likelihood that each base in a 2,000 nt section surrounding the FSE would pair with a base in the FSE (SFIG). Based on these predicted interactions, we selected ten coronaviruses – at least one from each genus (SFIG) – including SARS-CoV-2 as a positive control. Within the genus *Betacoronavirus*, we included all three SARS-related viruses – SARS coronaviruses 1 (`NC_004718.3`) and 2 (`NC_045512.2`) and bat coronavirus BM48-31 (`NC_014470.1`) – because they clustered into their own structural outgroup. The other three strains of *Betacoronavirus* that we selected were MERS coronavirus (`NC_019843.3`) with a predicted interaction at positions 510-530; and human coronavirus OC43 (`NC_006213.1`) and murine hepatitis virus strain A59 (`NC_048217.1`), both with a predicted upstream interaction at positions 10-20. We selected two strains of *Alphacoronavirus*: transmissible gastroenteritis virus (`NC_038861.1`) and bat coronavirus 1A (`NC_010437.1`), predicted to have interactions at positions 440-460 and 350-360, respectively. Avian infectious bronchitis virus strain

Beaudette (NC_001451.1) – a strain of *Gammacoronavirus* – was predicted to have a strong interaction at positions 330-350, while common moorhen coronavirus HKU21 (NC_016996.1) was the species of *Deltacoronavirus* with the most promising FSE interactions.

We reasoned that if an FSE does interact with a distant RNA element, then removing that element by truncating the RNA would break the interaction, causing a structural change in the FSE that could be detected though chemical probing. For each of the ten coronaviruses that passed the computational screen, we *in vitro* transcribed and performed DMS-MaPseq [13] on both a 239 nt segment comprising the FSE and minimal flanking sequences and a 1,799 nt segment encompassing the FSE and all sites with which it was predicted to interact. All coronaviruses except for human coronavirus OC43 and MERS coronavirus showed differences in their DMS reactivity profiles between the 239 nt and 1,799 nt segments (SFIG), suggesting long-range interactions involving the FSE.

SEARCH-MaP reveals long-range interactions involving the FSE in four other coronaviruses

To determine which RNA elements the FSE base-pairs with in each coronavirus, we performed SEARCH-MaP on the 1,799 nt RNA segment using DNA ASOs targeting the vicinity of the FSE (Figure 5). The rolling Spearman correlation coefficient (SCC) between the +ASO and no-ASO mutational profiles dipped below 0.9 at the ASO target site in every coronavirus segment, confirming the ASOs bound and altered the structure.
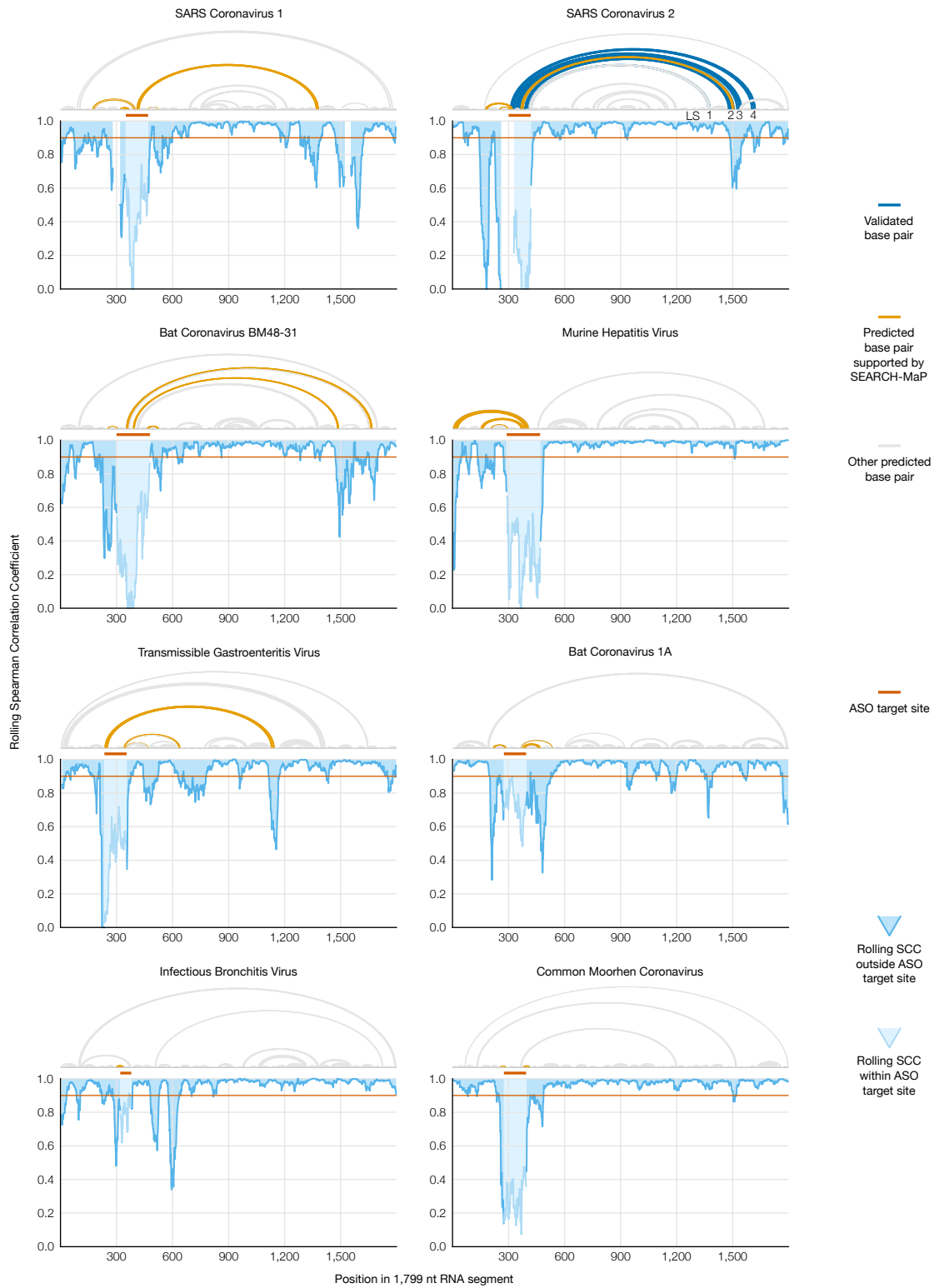
Figure 5:

To confirm we could detect long-range interactions, we compared the rolling SCC for the SARS-CoV-2 segment to our refined model of the long-range interaction (Figure 5, blue). The SCC dipped below 0.9 at positions 1,483-1,560 and at 1,611-1,642, which coincide with stems LS2-LS3 (positions 1,476-1,550 within the 1,799 nt segment) and

stem LS4 (positions 1,600-1,622). These dips were the two largest downstream of the FSE; although others (corresponding to no known base pairs) existed, they were barely below 0.9 and could have resulted from base pairing between these regions and other (non-FSE) regions. Near LS1 (positions 1,367-1,381), the SCC dipped only slightly to a minimum of 0.95, presumably because LS1 is the smallest (15 nt) and most isolated long-range stem. Therefore, this method was sensitive enough to detect all but the smallest long-range stem, and specific enough that the two largest dips corresponded to validated base pairs.

We found similar long-range interactions in SARS-CoV-1 and another SARS-related virus, Bat coronavirus BM48-31. Both viruses showed dips in SCC at roughly the same positions as LS2a-LS4 in SARS-CoV-2, indicating that they have homologous structures. SARS-CoV-1 also had a wide dip below 0.9 at positions 1,284-1,394, corresponding to a homologous LS1. Thus, three SARS-related viruses share this multi-stemmed long-range interaction involving the FSE, hinting this structure is functional.

In every other species except common moorhen coronavirus, we found prominent dips in SCC at least 200 nt from the ASO target site. To model potential base pairing between these dip positions and the FSE, we used the Fold program from RNAstructure [26] with the no-ASO ensemble average mutational profiles as DMS constraints [25]. Predictions based on ensemble averages (of all structural states) generally underperform those based on clustered mutational profiles (of fewer states or one state); the ensemble average prediction for SARS-CoV-2 included LS1 and LS2b but missed the other long-range stems. Nevertheless, we found long-range base pairs consistent with the SEARCH-MaP data for both murine hepatitis virus and transmissible gastroenteritis virus (Figure 5, orange). We conclude that long-range interactions involving the FSE occur more widely than in just SARS-CoV-2, including in *Alphacoronavirus* species.

We To verify that the long-range interaction also forms in live transmissible gastroenteritis virus (TGEV), we treated TGEV-infected ST cells with DMS (two biological replicates) and performed DMS-MaPseq (two technical replicates per biological replicate). The DMS reactivities were highly reproducible over the whole TGEV genome (r = 0.96-

0.97, SFIG). As expected, they differed from those of the 1.8 kb segment *in vitro* (r = 0.82, SFIG), showing why it is necessary to verify the long-range interaction in TGEV-infected cells.

First, to determine whether the FSE and the region with which it may interact form alternative structures, we amplified and deeply sequenced these two regions from each sample. Clustering the reads using SEISMIC-RNA revealed that both regions adopt at least two alternative structures. The two clusters of the downstream region differed most around positions 1,120-1,140 – the site of the 3' end of the predicted long-range interaction. In cluster 1 (63% of the ensemble), bases 1,129-1,136 (all part of the predicted interaction) had DMS reactivities less than 0.01; while in cluster 2, the DMS reactivities were all greater than 0.01. This result suggests that cluster 1 corresponds to the state in which the long-range interaction forms.

Discussion

In this work, we developed SEARCH-MaP and SEISMIC-RNA and applied them jointly to detect structural ensembles involving long-range RNA:RNA interactions in SARS-CoV-2 and other coronaviruses. This study is certainly not the first to perturb RNA structure with ASOs, nor even the first to use DMS-MaPseq to quantify the structural changes upon binding ASOs to SARS-CoV-2 RNA [30]. But while this previous study examined local structural perturbations caused by binding an ASO, we show that we can detect changes in the structure at more distant locations in an RNA molecule that interact with the nucleotides bound by an ASO.

A previous study detected two long-range RNA–RNA interactions in the genome of satellite tobacco mosaic virus by binding an ASO (in this case, an LNA 9-mer) to each site, followed by chemical probing [31]. However, SEARCH-MaP and SEISMIC-RNA go further by also determining the mutational profile and proportion of the interaction-formed and -unformed states (Figure 3c, d). With a collection of candidate structure models, these methods even reveal how adding an ASO ablates specific structures, collapsing the ensemble into one predominant structure (Figure 4d).

Many methods have been developed to find long-range (and intermolecular) RNA–RNA base pairing using crosslinking (with psoralen or a derivative), proximity ligation, and

deep sequencing [32, 33, 34, 11]. These methods require no prior knowledge of RNA–RNA interactions and have no limit to the length of the interactions they can detect. They do, however, suffer from several limitations including inefficient ligation [QUANTIFY, I think I read that less than 5% of molecules actually ligate] necessitating either enrichment or very deep sequencing, as well as bias towards U-rich sequences. They are not single-molecule techniques, either, meaning that although they can detect mutually exclusive base pairs, they cannot determine which specific alternative structures exist or quantify their proportions, as SEARCH-MaP/SEISMIC-RNA can. There is also no straightforward way to focus on one specific RNA–RNA interaction.

Another method based on applying many ASO "patches" in parallel and reading out the signal with microarray probes has also recently been developed [35]. Like proximity ligation, this method has no limitation to the length of the interactions it could find, yet it is also not a single-molecule technique, meaning that it cannot resolve individual structures in an ensemble.

SEARCH-MaP bears conceptual similarity to another method, mutate-and-map read out through next-generation sequencing (M2-seq) [36]. Both involve perturbing one region of an RNA molecule (in the case of M2-seq, by pre-installing mutations through error-prone PCR) and measuring the effects on other bases in the RNA using chemical probing. The major differences are the precision and scale of the interactions identified, as well as the throughput. M2-seq can pinpoint interactions down to the resolution of a single base pair, and is thus more precise than SEARCH-MaP. However, DMS-guided RNA structure prediction can propose structure models at single-base-pair resolution, which SEARCH-MaP can validate, and in this way achieve single-base-pair resolution. SEARCH-MaP is also capable of finding interactions over a much longer range because M2-seq requires the interacting bases to be in the same Illumina sequencing read. Within this length limit, one M2-seq experiment can theoretically find all pairwise interactions between bases, while one SEARCH-MaP experiment can find only interactions that involve the region to which the ASOs were hybridized. M2-seq is also limited by the formation of alternative structures. Some methods, such as [CITE something by Rhiju, maybe REEFIT] and DANCE-MaP [37], have been designed to work around this limitation SEARCH-MaP; however, [something by Rhiju] has [this problem], and DANCE-MaP requires extremely high sequencing depth of several million reads [MORE PRECISE].

SEARCH-MaP, by contrast, assumes from the start that the RNA may form alternative structures; for simply detecting long-range interactions, even a 5,000 read depth is sufficient coverage; and for clustering, we have found [SOME LIMIT].

Another limitation of SEARCH-MaP as presented here is that it cannot distinguish between direct and indirect interactions. If RNA segment A interacts with segment B, while B interacts with both segment A and C, then hybridizing an ASO to segment A would perturb the structure of B, which could consequentially perturb the structure of C. Hence, C would appear to interact with A, even though this interaction is indirect, through B. One possible workaround (not shown in this study) would be to mutate or hybridize an ASO to segment B, and then repeat the experiment with hybridizing an ASO to segment A. If the interaction between A and C is direct, then C should still be perturbed even when segment B is incapable of interacting with A or C. But if B mediates an indirect interaction between A and C, then disrupting B should eliminate the apparent interaction between A and C.

Functional long-range interactions up to four kilobases involving an FSE have been been found previously in two plant viruses [38, 39]. In both cases, frameshifting required the long-range interaction, suggesting that this interaction enables negative feedback on synthesis of viral RNA polymerase [38]. When polymerase levels are low, the interaction would form and stimulate frameshifting, which is needed to synthesize RNA polymerase. Once the polymerase had accumulated, it would begin to replicate the genomic RNA; in its passage from the genomic 3' end to the 5' end, it would disrupt the 3' side of the long-range interaction, attenuating frameshifting and reducing synthesis of more polymerase.

However, this strategy cannot be the role, if any, of the long-range interactions in coronaviruses. Unlike in the two plant viruses, a long-range interaction is not required to stimulate frameshifting in coronaviruses: numerous studies have shown that even the isolated FSE can cause 15 - 40% of ribosomes to frameshift [40, 41, 27, 10, 42, 43, 23]. In coronaviruses, the long-range interaction is not only unnecessary for frameshifting but also may even attenuate it, given that in SARS-CoV-2, the FSE-arch and the frameshift-stimulating pseudoknot seem to be mutually exclusive. Moreover, coronaviruses partition translation and RNA synthesis into two different cellular compartments (the cytosol and the double-membrane vesicles, respectively) [44], so structural changes induced by RNA polymerases would not be seen by ribosomes. If any of the long stems existed, they

would block the pseudoknot from forming, which suggests a mechanism by which the long-range interaction could regulate the structure – and possibly frameshifting activity – of the FSE. Although we had previously shown that AS1 overlaps and outcompetes PS1 [10], AS1 lies upstream of the slippery site and would be unwound by approaching ribosomes, while the long stems lie downstream and would not.

The functions of these long-range interactions involving the FSE in coronaviruses remain mysterious. However, given that they occur in multiple coronaviruses across at least two genera, it seems reasonable that they could play a role in the viral life cycle, possibly by affecting the rate of frameshifting. Further research may reveal new mechanisms of translational regulation in coronaviruses via long-range RNA:RNA interactions.

Methods

Screening coronavirus long-range interactions computationally

All coronaviruses with reference genomes in the NCBI Reference Sequence Database [29] were searched for using the following query:

```
refseq[filter] AND ("Alphacoronavirus"[Organism] OR
                    "Betacoronavirus"[Organism] OR
                    "Gammacoronavirus"[Organism] OR
                    "Deltacoronavirus"[Organism])
```

The complete record of every reference genome was downloaded both in FASTA format (for the reference sequence) and in Feature Table format (for feature locations). The location of the frameshift stimulating element (FSE) in each genome was estimated from the feature table, and the nearest instance of TTTAAAC was used as the slippery site, using a custom Python script. The 2,000 nt segment beginning 100 nt upstream of and ending 1,893 nt downstream of the slippery site was used for predicting long-range interactions involving the FSE. Genomes with ambiguous nucleotides (e.g. N) in this segment were discarded. For each coronavirus genome, up to 100 secondary structure models of the 2,000 nt segment were generated using Fold version 6.3 from RNAstructure [26]

with −M 100 and otherwise default parameters. Then, for each position, the fraction of models for the coronavirus in which the base at the position paired with any other base between positions 101 (the first base of the slippery sequence) and 250 was calculated using a custom Python script. The coronaviruses were clustered by their fraction vectors using the unweighted pair group method with arithmetic mean (UPGMA) and a euclidean distance metric, implemented in Seaborn version 0.11 [45] and SciPy version 1.7 [46]. The resulting hierarchically-clustered heatmap was examined manually to select coronaviruses based on the prominence of potential long-range interactions with the FSE (relatively large fractions far from positions 101-250).

References

[1] Ramya Rangan, Ivan N. Zheludev, Rachel J. Hagey, Edward A. Pham, Hannah K. Wayment-Steele, Jeffrey S. Glenn, and Rhiju Das. Rna genome conservation and secondary structure in sars-cov-2 and sars-related viruses: a first look. *RNA*, 26(8):937–959, 2020.

[2] Ilaria Manfredonia, Chandran Nithin, Almudena Ponce-Salvatierra, Pritha Ghosh, Tomasz K. Wirecki, Tycho Marinus, Natacha S. Ogando, Eric J. Snijder, Martijn J. van Hemert, Janusz M. Bujnicki, and Danny Incarnato. Genome-wide mapping of sars-cov-2 rna structures identifies therapeutically-relevant elements. *Nucleic Acids Research*, 48:12436–12452, 2020.

[3] Omer Ziv, Jonathan Price, Lyudmila Shalamova, Tsveta Kamenova, Ian Goodfellow, Friedemann Weber, and Eric A. Miska. The short- and long-range rna-rna interactome of sars-cov-2. *Molecular Cell*, 80:1067–1077.e5, 12 2020.

[4] Lei Sun, Pan Li, Xiaohui Ju, Jian Rao, Wenze Huang, Lili Ren, Shaojun Zhang, Tuanlin Xiong, Kui Xu, Xiaolin Zhou, Mingli Gong, Eric Miska, Qiang Ding, Jianwei Wang, and Qiangfeng Cliff Zhang. In vivo structural characterization of the sars-cov-2 rna genome identifies host proteins vulnerable to repurposed drugs. *Cell*, 184:1865–1883.e20, 2021.

[5] Yan Zhang, Kun Huang, Dejian Xie, Jian You Lau, Wenlong Shen, Ping Li, Dong Wang, Zhong Zou, Shu Shi, Hongguang Ren, Youliang Wang, Youzhi Mao, Meilin Jin, Grzegorz Kudla, and Zhihu Zhao. In vivo structure and dynamics of the sars-cov-2 rna genome. *Nature Communications*, 12:5695, 9 2021.

[6] Nicholas C. Huston, Han Wan, Madison S. Strine, Rafael de Cesaris Araujo Tavares, Craig B. Wilen, and Anna Marie Pyle. Comprehensive in vivo secondary structure of the sars-cov-2 genome reveals novel regulatory motifs and mechanisms. *Molecular Cell*, 81, 2021.

[7] Ramya Rangan, Andrew M. Watkins, Jose Chacon, Rachael Kretsch, Wipapat Kladwang, Ivan N. Zheludev, Jill Townley, Mats Rynge, Gregory Thain, and Rhiju Das. De novo 3d models of sars-cov-2 rna elements from consensus experimental secondary structures. *Nucleic Acids Research*, 49:3092–3108, 4 2021.

[8] Edoardo Morandi, Ilaria Manfredonia, Lisa M. Simon, Francesca Anselmi, Martijn J. van Hemert, Salvatore Oliviero, and Danny Incarnato. Genome-scale deconvolution of rna structure ensembles. *Nature Methods*, 18:249–252, 2 2021.

[9] Siwy Ling Yang, Louis DeFalco, Danielle E. Anderson, Yu Zhang, Jong Ghut Ashley Aw, Su Ying Lim, Xin Ni Lim, Kiat Yee Tan, Tong Zhang, Tanu Chawla, Yan Su, Alexander Lezhava, Andres Merits, Lin Fa Wang, Roland G. Huber, and Yue Wan. Comprehensive mapping of sars-cov-2 interactions in vivo reveals functional virus-host interactions. *Nature Communications*, 12, 2021.

[10] Tammy C.T. Lan, Matty F. Allan, Lauren E. Malsick, Jia Z. Woo, Chi Zhu, Fengrui Zhang, Stuti Khandwala, Sherry S.Y. Nyeo, Yu Sun, Junjie U. Guo, Mark Bathe, Anders Näär, Anthony Griffiths, and Silvi Rouskin. Secondary structural ensembles of the sars-cov-2 rna genome in infected cells. *Nature Communications*, 13:1128, 3 2022.

[11] Omer Ziv, Marta M. Gabryelska, Aaron T.L. Lun, Luca F.R. Gebert, Jessica Sheu-Gruttadauria, Luke W. Meredith, Zhong Yu Liu, Chun Kit Kwok, Cheng Feng Qin, Ian J. MacRae, Ian Goodfellow, John C. Marioni, Grzegorz Kudla, and Eric A. Miska. Comrades determines in vivo rna structures and interactions. *Nature Methods*, 15:785–788, 9 2018.

[12] Sharon Aviran and Danny Incarnato. Computational approaches for rna structure ensemble deconvolution from structure probing data. *Journal of Molecular Biology*, 434:167635, 9 2022.

[13] Meghan Zubradt, Paromita Gupta, Sitara Persad, Alan M. Lambowitz, Jonathan S. Weissman, and Silvi Rouskin. Dms-mapseq for genome-wide or targeted rna structure probing in vivo. *Nature Methods*, 2254:219–238, 2016.

[14] Nathan A. Siegfried, Steven Busan, Greggory M. Rice, Julie A.E. Nelson, and Kevin M. Weeks. Rna motif discovery by shape and mutational profiling (shape-map). *Nature methods*, 2014.

[15] Chringma Sherpa, Jason W. Rausch, Stuart F.J. Le Grice, Marie Louise Hammarskjold, and David Rekosh. The hiv-1 rev response element (rre) adopts alternative conformations that promote different rates of virus replication. *Nucleic Acids Research*, 43:4676–4686, 3 2015.

[16] Anthony M. Mustoe, Nicole N. Lama, Patrick S. Irving, Samuel W. Olson, and Kevin M. Weeks. Rna base-pairing complexity in living cells visualized by correlated chemical probing. *Proceedings of the National Academy of Sciences of the United States of America*, 116:24574–24582, 11 2019.

[17] Beth L. Nicholson and K. Andrew White. Functional long-range rna–rna interactions in positive-strand rna viruses. *Nature Reviews Microbiology*, 12:493–504, 6 2014.

[18] Ewan P. Plant and Jonathan D. Dinman. The role of programmed-1 ribosomal frameshifting in coronavirus propagation, 2008.

[19] Matthew F. Allan, Amir Brivanlou, and Silvi Rouskin. Rna levers and switches controlling viral gene expression. *Trends in Biochemical Sciences*, 48, 2023.

[20] Ian Brierley, Paul Digard, and Stephen C. Inglis. Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an rna pseudoknot. *Cell*, 1989.

[21] J. Herald and S. G. Siddell. An 'elaborated' pseudoknot is required for high frequency frameshifting during translation of hcv 229e polymerase mrna. *Nucleic Acids Research*, 21:5838–5842, 1993.

[22] Ewan P. Plant, Gabriela C. Pérez-Alvarado, Jonathan L. Jacobs, Bani Mukhopadhyay, Mirko Hennig, and Jonathan D. Dinman. A three-stemmed mrna pseudoknot in the sars coronavirus frameshift signal. *PLoS Biology*, 3:e172, 2005.

[23] Kaiming Zhang, Ivan N. Zheludev, Rachel J. Hagey, Raphael Haslecker, Yixuan J. Hou, Rachael Kretsch, Grigore D. Pintilie, Ramya Rangan, Wipapat Kladwang, Shanshan Li, Marie Teng Pei Wu, Edward A. Pham, Claire Bernardin-Souibgui, Ralph S. Baric, Timothy P. Sheahan, Victoria D'Souza, Jeffrey S. Glenn, Wah Chiu, and Rhiju Das. Cryo-em and antisense targeting of the 28-kda frameshift stimulation element from the sars-cov-2 rna genome. *Nature Structural & Molecular Biology*, 28:747–754, 8 2021.

[24] Christopher P. Jones and Adrian R. Ferré-D'Amaré. Crystal structure of the severe acute respiratory syndrome coronavirus 2 (sars-cov-2) frameshifting pseudoknot. *RNA*, 28:239–249, 2022.

[25] Pablo Cordero, Wipapat Kladwang, Christopher C. Vanlang, and Rhiju Das. Quantitative dimethyl sulfate mapping for automated rna secondary structure inference. *Biochemistry*, 51:7037–7039, 9 2012.

[26] David H. Mathews, Matthew D. Disney, Jessica L. Childs, Susan J. Schroeder, Michael Zuker, and Douglas H. Turner. Incorporating chemical modification constraints into a dynamic programming algorithm for prediction of rna secondary structure. *Proceedings of the National Academy of Sciences*, 101:7287–7292, 5 2004.

[27] Jamie A. Kelly, Alexandra N. Olson, Krishna Neupane, Sneha Munshi, Josue San Emeterio, Lois Pollack, Michael T. Woodside, and Jonathan D. Dinman. Structural and functional conservation of the programmed -1 ribosomal frameshift signal of sars coronavirus 2 (sars-cov-2). *Journal of Biological Chemistry*, 295:10741–10748, 7 2020.

[28] Mei Chi Su, Chung Te Chang, Chiu Hui Chu, Ching Hsiu Tsai, and Kung Yao Chang. An atypical rna pseudoknot stimulator and an upstream attenuation signal for -1 ribosomal frameshifting of sars coronavirus. *Nucleic Acids Research*, 2005.

[29] Nuala A. O'Leary, Mathew W. Wright, J. Rodney Brister, Stacy Ciufo, Diana Haddad, Rich McVeigh, Bhanu Rajput, Barbara Robbertse, Brian Smith-White, Danso Ako-Adjei, Alexander Astashyn, Azat Badretdin, Yiming Bao, Olga Blinkova, Vyacheslav Brover, Vyacheslav Chetvernin, Jinna Choi, Eric Cox, Olga Ermolaeva, Catherine M. Farrell, Tamara Goldfarb, Tripti Gupta, Daniel Haft, Eneida Hatcher, Wratko Hlavina, Vinita S. Joardar, Vamsi K. Kodali, Wenjun Li, Donna Maglott, Patrick Masterson, Kelly M. McGarvey, Michael R. Murphy, Kathleen O'Neill, Shashikant Pujar, Sanjida H. Rangwala, Daniel Rausch, Lillian D. Riddick, Conrad Schoch, Andrei Shkeda, Susan S. Storz, Hanzhen Sun, Francoise Thibaud-Nissen, Igor Tolstoy, Raymond E. Tully, Anjana R. Vatsan, Craig Wallin, David Webb, Wendy Wu, Melissa J. Landrum, Avi Kimchi, Tatiana Tatusova, Michael DiCuccio, Paul Kitts, Terence D. Murphy, Kim D. Pruitt, O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad Haft D, McVeigh R, Robbertse Rajput B, Robbertse Rajput B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb Gupta T, Goldfarb Gupta T, Haddad Haft D,

Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, Mc-Garvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A, Tatusova T, DiCuccio M, Kitts P, Murphy TD, and Pruitt KD. Reference sequence (refseq) database at ncbi: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, 44:D733–D745, 2016.

[30] Chi Zhu, Justin Y. Lee, Jia Z. Woo, Lei Xu, Xammy Nguyenla, Livia H. Yamashiro, Fei Ji, Scott B. Biering, Erik Van Dis, Federico Gonzalez, Douglas Fox, Eddie Wehri, Arjun Rustagi, Benjamin A. Pinsky, Julia Schaletzky, Catherine A. Blish, Charles Chiu, Eva Harris, Ruslan I. Sadreyev, Sarah Stanley, Sakari Kauppinen, Silvi Rouskin, and Anders M. Näär. An intranasal aso therapeutic targeting sars-cov-2. *Nature Communications*, 13:4503, 12 2022.

[31] Eva J. Archer, Mark A. Simpson, Nicholas J. Watts, Rory O'Kane, Bangchen Wang, Dorothy A. Erie, Alex McPherson, and Kevin M. Weeks. Long-range architecture in a viral rna genome. *Biochemistry*, 52(18):3182–3190, 2013. PMID: 23614526.

[32] Jong Ghut Ashley Aw, Yang Shen, Andreas Wilm, Miao Sun, Xin Ni Lim, Kum Loong Boon, Sidika Tapsin, Yun Shen Chan, Cheng Peow Tan, Adelene Y.L. Sim, Tong Zhang, Teodorus Theo Susanto, Zhiyan Fu, Niranjan Nagarajan, and Yue Wan. In vivo mapping of eukaryotic rna interactomes reveals principles of higher-order organization and regulation. *Molecular Cell*, 62:603–617, 2016.

[33] Zhipeng Lu, Qiangfeng Cliff Zhang, Byron Lee, Ryan A. Flynn, Martin A. Smith, James T. Robinson, Chen Davidovich, Anne R. Gooding, Karen J. Goodrich, John S. Mattick, Jill P. Mesirov, Thomas R. Cech, and Howard Y. Chang. Rna duplex map in living cells reveals higher-order transcriptome structure. *Cell*, 165:1267–1279, 2016.

[34] Eesha Sharma, Tim Sterne-Weiler, Dave O'Hanlon, and Benjamin J. Blencowe. Global mapping of human rna-rna interactions. *Molecular Cell*, 62:618–626, 2016.

[35] Timothy K Chiang, Ofer Kimchi, Herman K Dhaliwal, Daniel A Villarreal, Fernando F Vasquez, Vinothan N Manoharan, Michael P Brenner, and Rees F Garmann. Measuring intramolecular connectivity in long rna molecules using two-dimensional dna patch-probe arrays. *bioRxiv*, pages 2023–03, 2023.

[36] Clarence Y. Cheng, Wipapat Kladwang, Joseph D. Yesselman, and Rhiju Das. Rna structure inference through chemical mapping after accidental or intentional mutations. *Proceedings of the National Academy of Sciences of the United States of America*, 114:9876–9881, 9 2017.

[37] Samuel W. Olson, Anne Marie W. Turner, J. Winston Arney, Irfana Saleem, Chase A. Weidmann, David M. Margolis, Kevin M. Weeks, and Anthony M. Mustoe. Discovery of a large-scale, cell-state-responsive allosteric switch in the 7sk rna using dance-map. *Molecular Cell*, 82, 2022.

[38] Jennifer K. Barry and W. Allen Miller. A -1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral rna. *Proceedings of the National Academy of Sciences of the United States of America*, 99:11133–11138, 8 2002.

[39] Yuri Tajima, Hiro oki Iwakawa, Masanori Kaido, Kazuyuki Mise, and Tetsuro Okuno. A long-distance rna–rna interaction plays an important role in programmed - 1 ribosomal frameshfting in the translation of p88 replicase protein of red clover necrotic mosaic virus. *Virology*, 417:169–178, 8 2011.

[40] Pramod R. Bhatt, Alain Scaiola, Gary Loughran, Marc Leibundgut, Annika Kratzel, Romane Meurs, René Dreos, Kate M. O'Connor, Angus McMillan, Jeffrey W. Bode, Volker Thiel, David Gatfield, John F. Atkins, and Nenad Ban. Structural basis of ribosomal frameshifting during translation of the sars-cov-2 rna genome. *Science*, 372:1306–1313, 5 2021.

[41] Hafeez S. Haniff, Yuquan Tong, Xiaohui Liu, Jonathan L. Chen, Blessy M. Suresh, Ryan J. Andrews, Jake M. Peterson, Collin A. O'Leary, Raphael I. Benhamou, Walter N. Moss, and Matthew D. Disney. Targeting the sars-cov-2 rna genome with small molecule binders and ribonuclease targeting chimera (ribotac) degraders. *ACS Central Science*, 6:1713–1721, 2020.

[42] Ewan P. Plant, Rasa Rakauskaitė, Deborah R. Taylor, and Jonathan D. Dinman. Achieving a golden mean: Mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *Journal of Virology*, 84:4330–4340, 2010.

[43] Yu Sun, Laura Abriola, Rachel O. Niederer, Savannah F. Pedersen, Mia M. Alfajaro, Valter Silva Monteiro, Craig B. Wilen, Ya-Chi Ho, Wendy V. Gilbert, Yulia V. Surovtseva, Brett D. Lindenbach, and Junjie U. Guo. Restriction of sars-cov-2 replication by targeting programmed -1 ribosomal frameshifting. *Proceedings of the National Academy of Sciences of the United States of America*, 118:e2023051118, 6 2021.

[44] Georg Wolff, Charlotte E. Melia, Eric J. Snijder, and Montserrat Bárcena. Double-membrane vesicles as platforms for viral replication. *Trends in Microbiology*, 28:1022–1033, 12 2020.

[45] Michael Waskom. seaborn: statistical data visualization. *Journal of Open Source Software*, 6, 2021.

[46] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, 2020.