To knot and not: Multiple conformations of the SARS-CoV-2 frameshifting RNA element

Tamar Schlick,[*,†,‡,§] Qiyao Zhu,[‡] Abhishek Dey,[¶] Swati Jain,[†] Shuting Yan,[†] and Alain Laederach[*,¶]

†*Department of Chemistry, 100 Washington Square East, Silver Building, New York University, New York, NY 10003 U.S.A.*
‡*Courant Institute of Mathematical Sciences, New York University, 251 Mercer St., New York, NY 10012 U.S.A.*
¶*Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599*
§*NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, P.R. China*

E-mail: schlick@nyu.edu; alain@unc.edu

Abstract

The SARS-CoV-2 frameshifting RNA element (FSE) is an excellent target for therapeutic intervention against Covid-19. This small gene element employs a shifting mechanism to pause and backtrack the ribosome during translation between Open Reading Frames 1a and 1b, which code for viral polyproteins. Any interference with this process has profound effect on viral replication and propagation. Pinpointing the structures adapted by the FSE and associated structural transformations involved in frameshifting has been a challenge. Using our graph-theory-based modeling tools for representing RNA secondary structures, "RAG" (RNA-As-Graphs), and chemical structure probing experiments, we show that the 3-stem H-type pseudoknot (3_6 dual graph), long assumed to be the dominant structure, is replaced by a different, HL-type 3-stem pseudoknot (3_3) as more residues, in particular the slippery site's 7 residues, are considered. In addition, an unknotted 3-way junction RNA (3_5) emerges as a minor conformation. These three conformations share Stems 1 and 3, while the different Stem 2 may be involved in a conformational switch and possibly associations with the ribosome during translation. These structural and mechanistic insights advance our understanding of the SARS-CoV-2 frameshifting process and concomitant virus life cycle, and point to three avenues of therapeutic intervention.

Introduction

While the novel coronavirus agent, SARS-CoV-2, has decimated world economies, influenced political leadership, infected more than 121 million people, and claimed the lives of over 2.6 million, the scientific advances we have witnessed this past year are nothing short of miraculous. Besides numerous concerted vaccine development efforts that already bore fruit in warp speed, scientists across numerous disciplines have shifted research gears to aid in understanding the spread of the infection, determining optimal protocols to protect ourselves with cloth coverings and physical distancing, developing rapid and inexpensive tests for Covid-19, re-purposing known viral drugs to combat the disease, inventing new drugs, and unraveling the complex and multifarious biophysical aspects of the virus life cycle and infection trajectory. This amassed knowledge along with related prior research on RNAs and proteins has been instrumental to the success of the current mRNA-based vaccines by Moderna and Pfizer/BioNTECH, among the first to enter the clinic, and others that followed.

Through advanced molecular imaging and modeling strategies, scientific findings on Covid-19 and its SARS-CoV-2 agent have been pouring in. For example, outstanding contributions have been reported from cryo-electron microscopy imaging of the virus,[1] spike viral protein microscopy views[2] and motions,[3] RNA viral imaging scans and folding,[4,5] protein/ribosome views and motions,[6,7] and *in silico* drug screening studies.[8,9] This cumulative and rapid progress has helped us describe how the virus hijacks our own protein-synthesis machinery into making viral proteins efficiently and propose new lines of defense against the deadly disease it carries. These insights about the complex life cycle of the virus and mode of action are also invaluable for further development of drugs and other strategies to combat future viral epidemics that inevitably will arise. It is likely that this remarkable momentum of research will continue, even as massive vaccination programs help us approach normalcy in the coming months, to overcome our largely-stalled lives during 2020.

Although viral proteins have been a focus of many scientific groups, investigations of the RNA viral agent itself are crucial for understanding how the RNA invader replicates itself, is translated by the human ribosomal machinery, assembles, and synthesizes a suite of viral proteins that enable the continuation of its invasive trajectory. RNA-targeting therapeutics and vac-

cines can disarm the origin of the infection rather than its products and be more effective in the long term. Indeed, RNAs are excellent therapeutic targets due to the high conservation of the coronavirus viral genome and its close evolutionary relationship to other RNA viruses. However, the complexity of the RNA molecule and the lagging science about its modeling, imaging, and drug screening compared to proteins pose challenges. With technological improvements in RNA delivery systems, the rise of CRISPR-based gene editing systems,[10] and improved RNA modeling techniques,[11,12] this RNA focus is not only warranted but clearly successful, as evident by recent vaccines.

Of particular interest by many groups is the RNA frameshifting element (FSE), a small (<100 nucleotides) region in the open reading frame ORF1a,b region of the viral genome that codes for the polyproteins that initiate the cascade of viral protein synthesis. This FSE is attractive for several reasons. Its small size makes it highly amenable to modeling, since structure prediction tools and conformational sampling approaches are more reliable and less computationally intensive than for large RNAs. Likewise, techniques such as Selective 2′-Hydroxyl Acylation by Primer Extension (SHAPE) structure probing experiments[13,14] are relatively inexpensive and can be incorporated into RNA thermodynamic modeling approaches. Biologically, the FSE is responsible for the crucial −1 programmed ribosomal shifting mechanism utilized by many viruses to handle protein synthesis from overlapping reading frames.[15,16] By utilizing specific structural modules and associated conformational transitions, the FSE causes the ribosome to pause when encountering the ORF1b start site, which is out of register with respect to ORF1a. The ribosome then shifts the register by one base in the 5′ direction to define a different sequence of codons. Specifically, the −1 frameshifting alters the transcript UUU-UU*A(Leu)-AAC(Asn)-GGG at the second codon (asterisk) to back step by one amino acid and start as AAA-CGG(Arg) instead, so that translation resumes at CGG.

This crucial switch in the trajectory of viral protein synthesis offers an excellent target for therapeutic intervention, because mutation of key residues of the FSE and/or its binding to small molecules can interfere with this backtracking process. Indeed, several groups have explored how small-molecule agents such as *1,4-diazepane derivative 10* (MTDB) can inhibit this frameshifting process for both SARS-CoV and SARS-CoV-2.[17–19]

A structural focus on the FSE is key to defining target residues for drug binding as well as gene editing.[20] The FSE conformational landscape is particularly important because frameshifting efficiency does not appear to be a simple function of structural changes.[19,21] Thus, complex interactions and mechanisms are likely involved both within the FSE and between the FSE and the ribosome during protein synthesis.

Here we focus on better understanding of this FSE conformational landscape using a combination of complementary graph-based modeling and chemical reactivity experiments. The FSE occupies 84 residues between 13462–13545 in the SARS-CoV-2 RNA genome of 29891nt, and consists of a 7-residue slippery site (UUUAAAC) and a 77-residue RNA (Fig. 1). Already, several groups have explored FSE structure by modeling, *in vivo* SHAPE chemical reactivity, NMR, Cryo-EM, and other biophysical mutational profiling and scanning experiments.[4,22–24] Many have characterized the FSE as a 3-stem H-type pseudoknot with colinear Stems 1 and 2 intertwined via a pseudoknot and perpendicular Stem 3. This association has persisted from the SARS-CoV FSE characterization,[19] which differs in only one base from the SARS-CoV-2 FSE (residue A13533 in Covid-19 is C in SARS, Fig. 1). However, depending on the modeling software and experimental technique, proposals of alternative structures have been offered for SARS-CoV-2, both pseudoknotted and unknotted (see below). Yet no coherent picture of this conformational landscape has emerged. Because of the crucial relationship of the FSE structure and the frameshifting mechanism, it is important to unravel the FSE conformational landscape.

In our prior work[20] (see also commentary[25]), we defined target residues for drug binding and gene editing of the FSE from designed minimal mutants that dramatically transform the conformation of the FSE. Our RAG (RNA-As-Graphs) machinery represents RNA 2D structure as coarse-grained dual graphs, where double-stranded RNA helices are represented as vertices, and loop strands are edges. The common H-type pseudoknotted structure of the FSE corresponds to the 3_6 dual graph as shown in Fig. 1. Using our RAG-based genetic algorithms for RNA design by inverse folding,[26] we designed mutants that altered the 3_6 conformation for a 77nt FSE (no slippery site) (Fig. 1) into 3-stem and 2-stem structures with and without pseudoknots. Our designs showed that mere single and double mutants can cause drastic changes in the RNA secondary structure, such as deleting Stem 2, or breaking the pseudoknot between Stems 1 and 2. Microsecond molecular dynamics simulations of these mutants modeled at atomistic detail with explicit solvent also demonstrated the stability of these alternative forms and highlighted the flexibility of Stem 2 of the FSE. Among these four mutants we explored, a 3-way junction (dual graph 3_5) is further investigated here. Note that coarse-grained graph representations are robust and capture the topology of the RNA while allowing for differences in the lengths of stems and loops; thus, the same graph corresponds to multiple 2D models that differ in sizes of stems and loops. This coarser representation of RNAs was key in deducing the above FSE mutations efficiently.

We also noted that structural predictions of the FSE by available programs depend on the length considered,[20] in particular, whether the slippery site is included. This inclusion is both computationally and biologically meaningful, since the slippery site is thought to be inaccessible while the FSE is in direct interaction

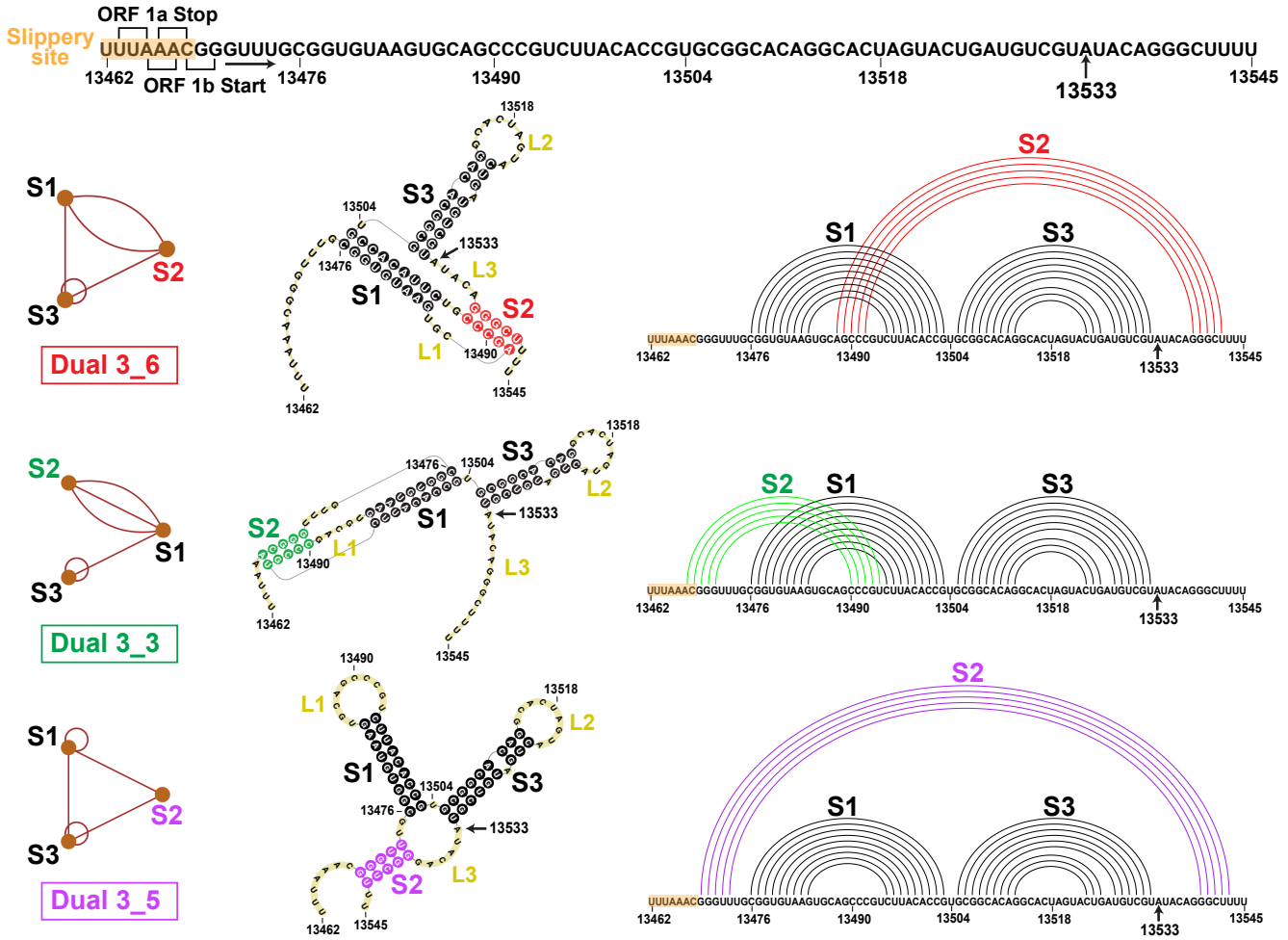


Figure 1: Three relevant 2D structures for the SARS-CoV-2 84nt frameshifting element (residues 13462–13545) emerging from our analysis in this work that combines 2D structure prediction, SHAPE structural probing, and thermodynamic ensemble modeling. At top is the 84nt sequence, with the 7nt slippery site highlighted and A13533 labeled (C in SARS). The ORF1a end and ORF1b start codons for the overlapping regions are marked. For each 2D structure, H-type 3_6 pseudoknot, HL-type 3_3 pseudoknot, and three-way junction 3_5 (unknotted RNA), corresponding dual graphs, 2D structures, and corresponding arc plots are shown, with color coded stems and loops labeled.

with the ribosome, but possibly free otherwise. Besides these 7 residues, neighboring units also influence the predicted topologies obtained. We showed that the sequence context of 77nt, 84nt, and 144nt leads to various structure predictions for the FSE that are both pseudoknotted and unknotted (Fig. 3 here too).[20]

Here, we begin to untangle this mystery by taking a closer look at the length-dependence structure of the FSE by graph theory modeling combined with chemical structure probing experiments designed to address the length context of the FSE. Our combined analysis leads us to describe a conformational landscape with *three viable structures* of the FSE: two pseudoknotted RNAs (3_6 and 3_3 in our dual graph notation, or H-type and HL-type 3-stem pseudoknots), and one unknotted, 3-way junction RNA (3_5) (Fig. 1).* Stems 1 and 2 in the 3_6 pseudoknot are coaxially stacked, with Stem 1 loop base pairing with the 3′ end of Stem 2, and Stem 3 is oriented perpendicular to the Stem 1/Stem 2 plane. In the 3_3 pseudoknot, Stem 1's loop binds the 5′ end of Stem 2 that includes slippery site residues. The flexible Stem 2 may thus be involved in a switch between these conformations and possibly associations with the ribosome during protein translation. It may also define a co-transcriptional kinetic folding trap. Together, the SHAPE data and statistical landscape modeling help describe the relation between FSE length and structure, as well as implications to frameshifting mechanisms involving the ribosome. These results help consolidate FSE reports to date and define new therapeutic avenues for regulating frameshifting efficiency and hence Covid-19 infections.

Results

Multiple sequence alignment and variant analysis of coronaviruses emphasize FSE features

To put into context the FSE structure of SARS-CoV-2 (see Fig. 1) and pinpoint the relative flexibility of the different stems, we analyze the sequence similarity of the FSE region in the coronavirus family by multiple sequence alignment. For 11 other coronaviruses with -1 programmed ribosomal frameshifting (PRF), we show the multiple sequence alignment for the FSE region in Fig. 2. Both the overall genome homology with SARS-CoV-2 and with only the FSE are indicated. The darker purple shadings indicate greater sequence homology. Based on our work presented in[20] and here, we annotate Stems 1 and 3 of SARS-CoV-2 FSE in black, and Stem 2 corresponding to the H-type pseudoknot (dual graph 3_6) in red, the HL-type pseudoknot (3_3) in green, and the 3-way junction (3_5) in purple.

Stem 1 is highly conserved, suggesting a consensus Stem 1 for all twelve coronaviruses. Stem 3 is longer than Stem 2, but its residues have undergone variations that are coronavirus specific, suggesting different locations and lengths for Stem 3 in different coronaviruses. Stem 2 is the shortest and the most flexible. The two pseudoknots 3_6 and 3_3 have equally strong Stems 2, both made of one AU and four GC base pairs. The two Stems 2 share the same central CCC region (residues 13490–13492) but involve different base pairing orientations as shown in Fig. 1. The middle C13491 in the shared region is mutated to U in five more distant coronaviruses. Interestingly, compensatory mutations from G to A occur at complementary locations for both 3_6 (residue 13539) and 3_3 (13470) Stem 2, lending support to both structures. For coronaviruses NL63, 229E, and PEDV, the covariation evidence is stronger for 3_3 Stem 2, suggesting a preference for these less homologous coronavirus FSEs. Stem 2 of the 3_5 junction appears less stable, made of one GC and four GU base pairs.

To demonstrate the sequence conservation of the SARS-CoV-2 FSE RNA, we analyze 459421 variants of SARS-CoV-2 deposited on GISAID (Global Initiative on Sharing All Influenza Data) database[27] by February 12, 2021 (Fig. S1). Only 8504 or 2% exhibit mutations in the FSE segment. Among the mutated sequences, 98% are single mutants. Mutation maps for the 84nt FSE and the spike gene segment per nucleotide (plotted on different scales) show that the spike gene region has an order of magnitude more mutations than the FSE. Interestingly, residue A13533, which is C in SARS-CoV, is never mutated to C, but only to G. Further analysis of recent highly transmissible British (B.1.1.7), South Africa (B.1.351), Brazil (P.1), and New York City (B.1.526) Covid-19 variants also show concentrated mutations in the spike gene region, with 4-12 residues having mutation rates $> 85\%$, and very few (random) mutations in the FSE (Fig. S2). This analysis reinforces the high conservation of the FSE region and its suitability for anti-viral therapy.

Length dependence RNA 2D structure predictions raise caveats

Several works have reported scanning approaches for RNAs where windows of 120nt are considered to predict secondary structural elements.[24,29,30] This length value is believed to yield reasonable predictions for RNAs.[31,32] To understand the folding and translation mechanism when the FSE is interacting with the ribosome, as well as interpret RNA structural predictions from computational programs, we further explore how structures of four experimentally-solved RNAs with pseudoknots are predicted for different RNA lengths.

We applied five 2D folding programs that can predict pseudoknots (PKNOTS,[33] NUPACK,[34] IPknot,[35] ProbKnot,[36] and vsfold5[37]) to four RNAs with pseudoknots (lysine riboswitch, PDB: 3DIO; lariat-capping ribozyme, PDB: 6GYV; glmS ribozyme, PDB: 2GCV; T-Box riboregulator, PDB: 6UFG), for three different lengths ("short", 120nt, and "all"). These four RNAs have lengths 151-192nt and they contain pseudoknots of similar sizes (79-106nt) to the FSE pseudoknot. Fig. S3 shows the accuracy of the predictions compared to the experimentally determined structure. Although we see that the accuracy varies with the method and RNA, the 120nt window recommended in the literature appears reasonable in general. This analysis suggests that consideration of only 77-84 residues of the FSE in computational structure prediction may miss features of the structures. Hence, we introduce a 144nt FSE construct, adding 30 upstream and 30 downstream nucleotides to the 84nt FSE.

Length-dependent predictions for the FSE were already presented in our prior work.[20] Here we show further predictions with associated graph topologies in Fig. 3 using PKNOTS, NUPACK, IPknot, and ProbKnot. In each case, 3 lengths (77, 84, and 144nt) are used, and the 2D arc plots for hydrogen bonding are shown, along with the associated dual graphs.

We see that for 77nt, 3 out of the 4 programs predict a 3_6 pseudoknot (H-type), but ProbKnot predicts the 3_5 3-way junction. For 84nt, only PKNOTS predicts the 3_6 pseudoknot, while ProbKnot predicts the 3_3 pseudoknot (HL-type), and IPknot and NUPACK predict a two-pseudoknot structure 4_7. This 4_7 graph can be partitioned, using our partition algorithm for dual graphs,[38] into subgraphs 3_3 and 2_3 (see Fig. S4), with the former corresponding to the 3_3 pseudoknot, and the latter to the new pseudoknot formed by the $3'$ end intertwining with Stem 3. For 144nt FSE, the predictions are quite different. Only Stem 1 and two upstream hairpins that block 3_3 Stem 2 are consistently predicted. Stem 3 appears in three predictions, while Stem 2 of 3_6 pseudoknot only arises in IPknot and NUPACK. The central 3_6 pseudoknot is the same by

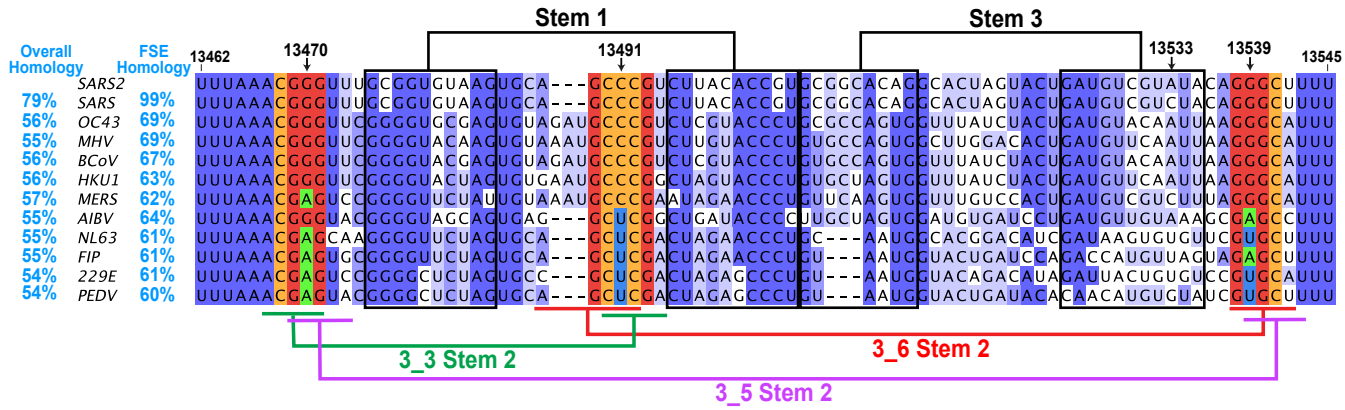


Figure 2: Multiple sequence alignments (MSA) of coronavirus frameshifting elements by ClustalW2.[28] The other 11 coronaviruses are aligned with SARS-CoV-2 84nt FSE region, with sequence similarities shown for both the whole genome and the FSE region (in decreasing order). Nucleotides are colored based on sequence conservation. Stems are marked based on our prior work[20] and analysis here: black for Stems 1 and 3, and red/green/purple for Stem 2 of 3_6/3_3/3_5, consistent with Fig. 1. The 4 GC base pairs in Stem 2 of 3_6 and 3_3 are colored by nucleotide identity. (Coronaviruses are: SARS2: SARS coronavirus 2; SARS: SARS coronavirus; OC43: human coronavirus OC43; MHV: murine hepatitis virus; BCoV: bovine coronavirus; HKU1: human coronavirus HKU1; MERS: middle east respiratory syndrome coronavirus; AIBV: avian infectious bronchitis virus; NL63: human coronavirus NL63; FIP: feline infectious peritonitis virus; 229E: human coronavirus 229E; PEDV: porcine epidemic diarrhea virus.)

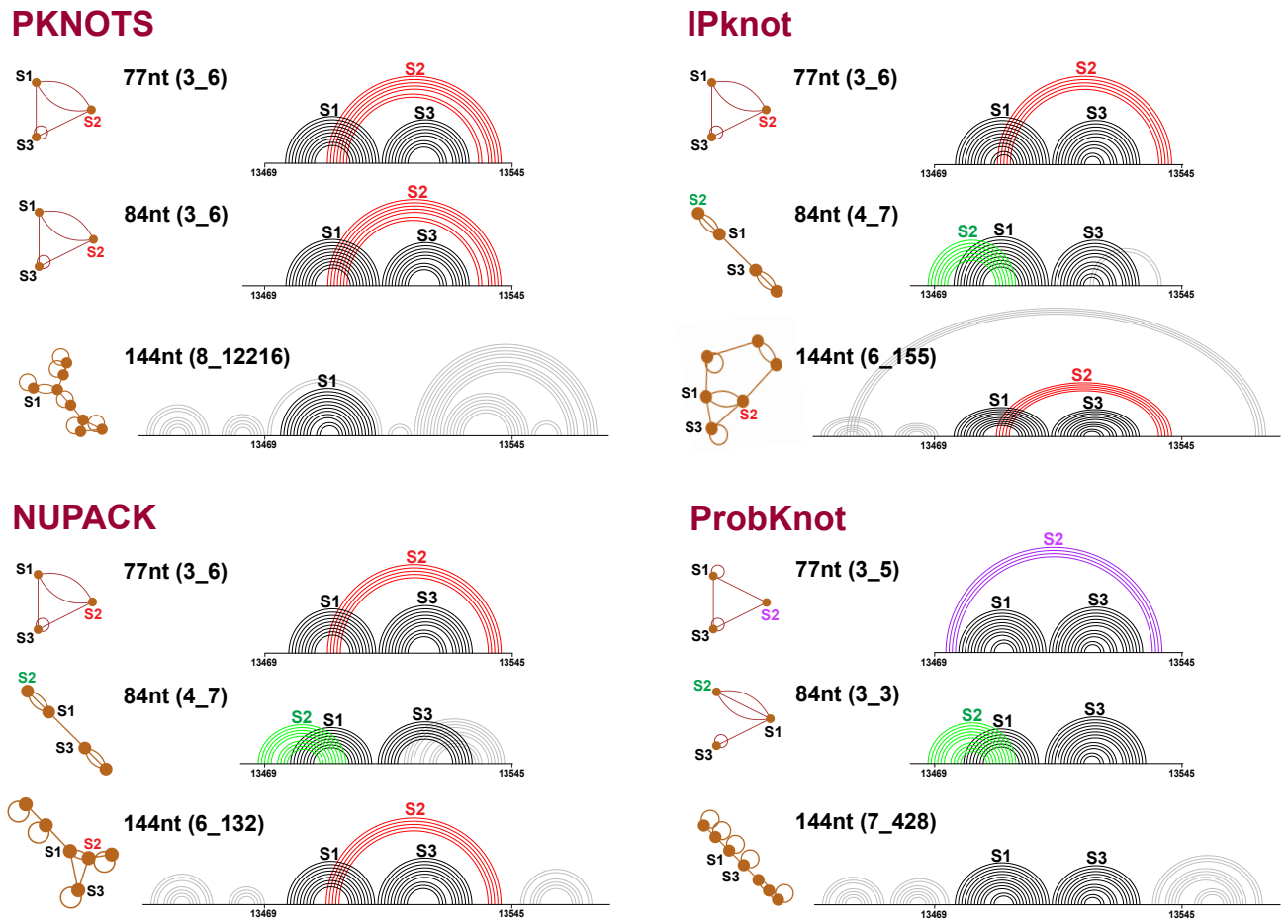


Figure 3: Predicted structures for the frameshifting element using PKNOTS, NUPACK, IPknot, and ProbKnot (see text). For each program, 3 different sequence lengths are used: 77, 84, and 144nt. The predicted structures (aligned with common 77nt subsequence, labeled) are shown as arc plots, with corresponding dual graphs listed. Stems 1 and 3 are colored black, and Stem 2 of 3_6, 3_3, and 3_5 are colored red, green, and purple, respectively, if present.

both programs, but IPknot binds the 3′ end with the 5′ end hairpin loop to form another pseudoknot (6_155) while NUPACK predicts a 3′ end hairpin (6_132). Both multiple sequence alignment and length-dependent predictions show that Stem 1 is highly conserved, while Stem 2 is flexible (Fig. 2).

SHAPE reactivity data reveals dominant alternative pseudoknot in longer sequence contexts and minor 3-way junction

To experimentally probe the formation of alternative structures in the SARS-CoV-2 FSE, we investigate the SHAPE reactivity of two RNA FSE constructs of 77nt (residues 13469–13545) and 144nt (residues 13432–13575) in Fig. 4.

In general, SHAPE (Selective 2′-Hydroxyl Acylation by Primer Extension) experiments provide structural anchors for interpreting RNA structures by exploiting the high reactivity of free 2′-hydroxyl segments of the RNA ribose sugar to suitable chemical reagents. The measured reactivities at each nucleotide are directly correlated to the local RNA flexibility, and the paired bases will generally have low reactivity. These experimental data are used to define modified base pair probabilities that guide the energy minimization in the structure prediction program (ShapeKnots).[39,40]

For each FSE length, we use two replicates, with 5NIA reagent and Bicine buffer (see Methods, SI for alignments of replicates, and Table 1). In Fig. 4, the SHAPE reactivities of Replicate 1 are shown as histograms plotted per residue, and the arc plots above correspond to the dominant prediction using ShapeKnots. The arc plots below the reactivity data correspond to minor conformers. Because ShapeKnots predicts multiple structures ranked by free energies, we apply Boltzmann weighting ($p_i = \exp(-E_i/(k_B\,\mathrm{T}))$ where p_i and E_i are the probability and free energy for conformer i, k_B is the Boltzmann constant, and T is room temperature, set at 37°C) to calculate the energy landscape contribution of each conformer.

Consistent with the modeling for 77nt, when we incorporate its SHAPE reactivity in ShapeKnots 2D structure ensemble predictions, we find that 98% of structures form the 3_6 pseudoknot, with the 3_5 3-way junction playing a minor role (Fig. 4A). The 3_6 pseudoknot has the same structure as the predictions by PKNOTS, IPknot and NUPACK (Fig. 3), with small variations in stem lengths. The 3_5 junction nevertheless has a shifted Stem 2 towards the 5′ and 3′ end, compared with the prediction by ProbKnot.

The same experiment on the 144nt construct detects the 3_6 pseudoknot only in 4.4% of the population, while the 3_3 pseudoknot represents 95.6% of the landscape (Fig. 4B). The 144nt 3_6 conformation agrees with NUPACK's prediction in Fig. 3 (dual graph 6_132). Comparing the two pseudoknotted structures (top versus bottom arc plots), we see that Stems 1 and 3 are very similar, but Stem 2 is different. In 3_3, the pseudoknot involves Stem 1 intertwining with Stem 2 at the 5′ end of the FSE, while for 3_6, Stem 1 loop region hydrogen bonds with the 3′ end of the FSE to form Stem 2 (Fig. 1).

The computed difference in SHAPE reactivity (77nt reactivity minus 144nt reactivity) reveals changes consistent with these findings in two key regions (Fig. 4C): Stem 1 with its loop and the 77nt 3′ end. In the Stem 1 loop, two residues A20 and G21 (numbered in the 77nt context, equivalent to A13488 and G13489) are only paired in the 3_6 Stem 2, and are more flexible in the 144nt system. Commensurately, the complementary residue U74 of A20 on the 3′ end also has increased flexibility in 144nt. Similarly, for 144nt, Stem S_F flanking the 3_3 pseudoknot involves a critical residue A69 (less flexible in 144nt) absent from base pairs associated with the 3_6 pseudoknot and 3_5 junction.

Replicate 2 (Fig. S5) reaffirms our finding of a dominant 3_6 pseudoknot (97%) and a minor 3_5 junction for 77nt. For 144nt, besides the 3_3 pseudoknot conformation (dual graph 7_2192) that was dominant in Replicate 1, Replicate 2 yields another 3_3-containing structure (dual 6_383). The two structures share the central 3_3 pseudoknot of Replicate 1 but differ in the flanking regions. Namely, in 57% of the conformations, S_F is replaced by a hairpin at the 3′ end (see Fig. S5). In Fig. S6A, we see that the two aligned replicates for each construct agree well with one another, especially for the 77nt construct.

We also used dimethyl sulfate (DMS) chemical probing coupled with mutational profiling to identify correlations in the structure. The PairMap technique identifies correlation[41] to suggest not only which residues are paired but with whom they may pair. In Fig. S6B, the dark arcs from PairMap indicate the principal interactions, while the lighter colored arcs correspond to minor interactions. We see that, consistent with the multiple sequence alignment (Fig. 2), Stems 1 and 3 (for 77nt) and Stem 1 (for 144nt) are strongly preserved, while Stem 2 is more tentative. The minor Stem 2 for both lengths corresponds precisely to the Stem 2 in the two pseudoknotted structures above. This additional experimental approach supports our findings and is based on a direct analysis of DMS mutational reactivities, independent of thermodynamic modeling.

Mutant predictions for dominating the conformational landscape by 3_6, 3_3, and 3_5 topologies are confirmed by SHAPE

Our SHAPE and correlated DMS experiments suggest three relevant structures that make up the FSE conformational landscape: two pseudoknots (3_6 and 3_3), and a 3-way junction (3_5). As this unusual conformational flexibility may play a mechanistic role in the frameshifting event, we sought to stabilize each conformer by minimal mutations. Such analysis also provides insights into

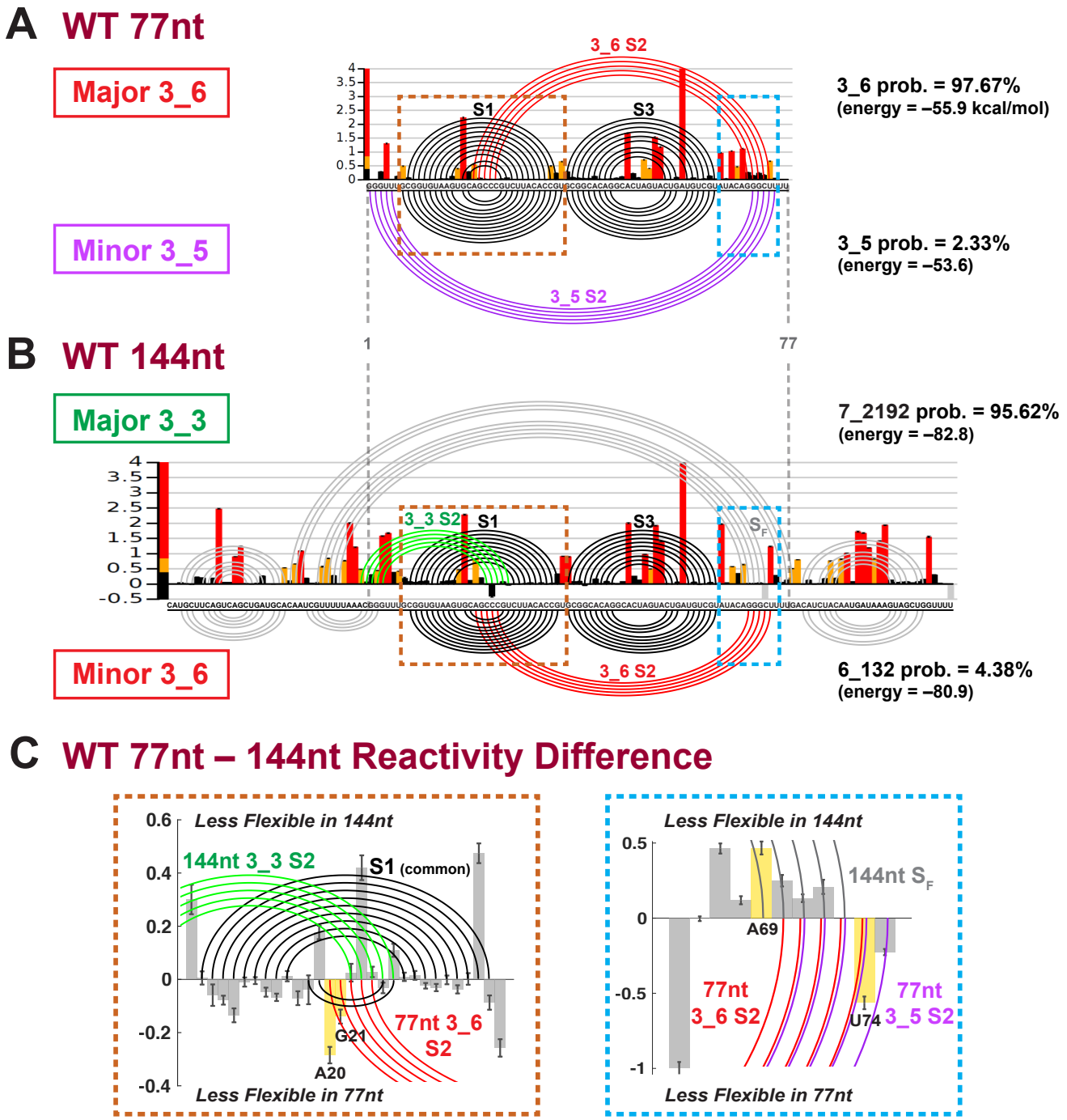


Figure 4: SHAPE reactivity analysis for SARS-CoV-2 frameshifting element for 77nt and 144nt (Replicate 1). (A) The SHAPE reactivity for the 77nt construct is plotted by bars, with red/yellow/black representing high/medium/low reactivity. The arc plot at top shows the dominant 3_6 pseudoknot predicted by ShapeKnots (98% of conformational space), and at bottom is the minor 3_5 (2% of landscape). Stems are labeled, and the Gibbs free energy ($kcal/mol$) and Boltzmann distribution probabilities are given. (B) SHAPE reactivity and ShapeKnots predictions for 144nt construct. (C) Reactivity differences between the two constructs are shown for two enlarged key regions highlighted in A,B, with positive/negative differences indicating less flexibility in the 144nt/77nt construct. Base pairs in the 144nt 3_3 conformation are plotted by arcs at top, and 77nt 3_6 or 3_5 at bottom. Critical residues for reactivity comparisons are highlighted.

possible transitions between the three conformers.

We apply our RAG-based software RAG-IF[26] as developed and applied in our prior work[20] to determine minimal mutations for each conformer to dominate the landscape. Briefly, our genetic algorithm works by transforming one dual graph into another by iterating

Table 1: Summary of ShapeKnots prediction results for wildtype frameshifting element and mutants developed and tested in this work. For each construct, the probability and free energy ($kcal/mol$) predicted for 3_6 pseudoknot, 3_5 junction and 3_3 pseudoknot are shown. The mutations are annotated by their positions in the relative constructs; 77nt construct covers residues 13469–13545, 87nt covers 13459–13545, 144nt covers 13432–13575. PSM: Pseudoknot-strengthening mutant; see next section.

Construct	Replicate	3_6 prob. (energy)	3_5	3_3
WT 77nt	1	97.67% (−55.9)	2.33% (−53.6)	None
	2	97.26% (−60.0)	2.74% (−57.8)	None
WT 87nt	1	0.13% (−60.1)	None	99.87% (−64.2)
	2	0.06% (−57.8)	None	99.94% (−62.4)
WT 144nt	1	4.38% (−80.9)	None	95.62% (−82.8)
	2	2.26% (−74.2)	None	97.74% (−76.2)
3_6 PSM 77nt [G3U, U4A, G18A, C19A, C68A, A69C]	1	100% (−63.1)	None	None
	2	100% (−51.6)	None	None
3_6 PSM 144nt [G40U, U41A, G55A, C56A, C105A, A106C, C137A]	1	100% (−101.0)	None	None
	2	100% (−106.2)	None	None
3_3 PSM 77nt [U4C, G71A, G72U]	1	None	None	100% (−59.7)
	2	None	None	100% (−61.0)
3_5 Mutant 77nt [G72C, U74C]	1	None	100% (−66.8)	None
	2	None	100% (−69.6)	None

on a sequence of mutations in pre-selected regions so as to minimize the difference (measured by Hamming distance) between the current and target graph, in the spirit of a natural selection process; the fold of each graph is determined by a consensus between two 2D folding programs. See Methods and Refs.[26]

To design the 3_6 pseudoknot-strengthening mutant (PSM), we apply RAG-IF to transform the 3_5 predicted by ProbKnot (Fig. 3) onto 3_6 (Fig. 5A). A 4-residue mutant [G3U, U4A, C68A, A69C], which breaks Stem 2 of 3_5 and creates two extra base pairs for Stem 2 of 3_6, is selected for the 77nt construct. With additional 2D prediction program screening (see Fig. S7A), we add two mutations [G18A, C19A] to further strengthen Stem 2 with 4 additional base pairs. For 144nt, after testing the above 6 mutations using four 2D prediction programs (Fig. S7B), we add a mutation to the 3′ end to inhibit a stem that interferes with Stem 2 of 3_6. The resulting 7-residue mutant is [G40U, U41A, G55A, C56A, C105A, A106C, C137A].

Subsequent SHAPE experiments confirm our predictions for both the 77nt and 144nt constructs of this 3_6 PSM: 100% of the landscape is now occupied by the 3_6 pseudoknot (Fig. 5A), when chemical reactivity data are incorporated in the 2D structure prediction by ShapeKnots. The 3_6 Stem 2 has 7 instead of the expected 9 base pairs, but is longer than the wildtype Stem 2 (Fig. 4A). Moreover, when we compare the 3_6 PSM with the wildtype for 77nt constructs (Fig. S8A), the reactivity differences support the two new base pairs in Stem 2. Residue G25 was base paired with C19 in the wildtype FSE, but after the C19A and A69C mutations, it pairs with 69C. As a result, C19 has increased flexibility and 69C shows decreased flexibility in the PSM SHAPE data. Similarly, for the new base pair U26 with 68A, we note decreased reactivity for 68A.

We also design a 3_3 pseudoknot-strengthening mutant for the 77nt FSE similarly. We choose a triple mutant [U4C, G71A, G72U] that is predicted to form 5-7 base pairs for 3_3 Stem 2 (see Fig. S9 and Methods). Subsequent SHAPE reactivities in Fig. 5B show that the conformational landscape is now 100% 3_3 pseudoknot. The comparison between the 3_3 PSM and the wildtype 77nt reactivities (Fig. S8B) show very small differences in the Stem 1 region, because Stem 2 of the wildtype 3_6 and the PSM 3_3 overlap. Nevertheless, we see increases in flexibility at the 3′ end, where the 3′ strand of Stem 2 in 3_6 and 3_5 locate, supporting 3_3 pseudoknot over the other two conformations in this predicted mutant.

Because the unknotted 3-way junction (3_5) also emerges as a minor player in the 77nt FSE conformational landscape, we obtain reactivity data for the double mutant we had predicted in our prior work[20] to stabilize this fold over the two pseudoknots. Fig. 5C shows that merely two mutations [G72C, U74C] on the 3′ edge of 3_6 Stem 2 accomplish this dramatic change. This 3-way junction becomes the sole conformer in the 77nt mutant landscape, compared to 2–3% in the wildtype (Fig. 4A). By examining reactivity differences with the wildtype 77nt (Fig. S8C), we find that residues in the loop region of Stem 1, which are based paired in the wildtype 3_6 Stem 2, become more flexible in this mutant. Moreover, a 3′ end residue A69, which is newly base paired in this mutant's Stem 2, has decreased reactivity, again supporting the 3_5 conformation.

In summary, the combined evidence points to a conformational landscape for the SARS-CoV-2 FSE that is sensitive to the sequence length and highlights two ma-

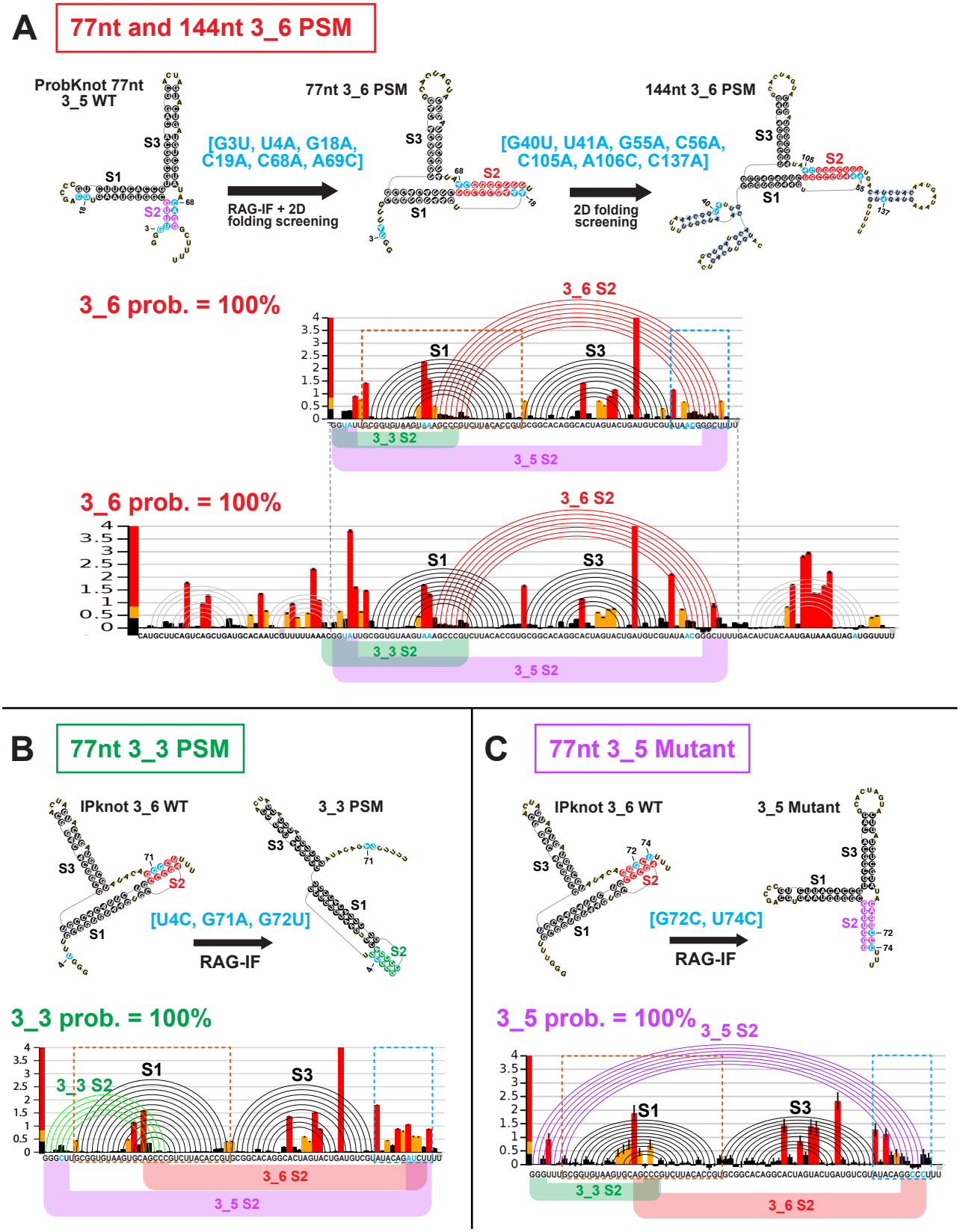


Figure 5: Design and SHAPE analysis for (A) 77nt and 144nt 3_6 pseudoknot-strengthening mutant (PSM), (B) 77nt 3_3 PSM, and (C) 77nt 3_5 Mutant. For each mutant, we show the design flow, where we use RAG-IF and multiple 2D structure prediction program screening to determine mutations that stabilize the 3_6 pseudoknot, 3_3 pseudoknot, or 3_5 junction. Mutations are highlighted in blue. The SHAPE reactivity bar plots, and arc plots of the structure predicted by ShapeKnots, with alternative Stem 2 positions are shown. See Fig. S8 for reactivity differences between the mutants and the wildtype for two boxed key regions.

jor players — an H-type pseudoknot (3_6 dual graph) and an HL-type pseudoknot (3_3 dual graph) — as well as a minor 3-way junction (unknotted) (3_5). Our mutant predictions are confirmed for strengthening all three structures, using two SHAPE replicates and their reactivity analyses (see Table 1 and Fig. S10 for alignments). Although the 3_6 pseudoknot (98%) dominates the wildtype 77nt landscape, our mutation experiments show that with only 2 or 3 mutations, we can shift the landscape to be 100% 3_5 3-way junction or 100% 3_3 pseudoknot. Hence, all 3 conformations are viable for the FSE, and they may not be too far away from one another from a sequence landscape point of view.

Consensus conformational landscape of the FSE clarifies length dependence

To consolidate the above information, we now estimate the energy landscape for the three viable conformations as a function of sequence length. We consider two methods of expanding the FSE sequence to explore this length dependence further and pinpoint possible conformational transitions: (1) asymmetrically adding residues only on the 5′ end from 77nt to 114nt (after adding the 7nt slippery site), and (2) symmetrically expanding both ends from 77nt to 144nt (after adding the 7nt slippery site). The asymmetric approach helps determine the shortest FSE length for obtaining the 3_3 pseudoknot (see SHAPE experiment below), and is also realistic for ribosomal interactions. The symmetric expansion helps interpret the full landscape.

For each length, we first extract experimental reactivities for corresponding residues from 144nt construct, and renormalize them to have the same mean value as the 77nt construct. Second, we predict the RNA 2D structures using ShapeKnots, along with Gibbs free energies, and then calculate respective Boltzmann probabilities. Third, we sum up probabilities for all structures containing independently folded 3_6 or 3_3 pseudoknot, and display populations in red (3_6) and green (3_3) for each length in Fig. 6. *Although the reactivities used here for different sequence lengths are not the real data from the folded RNA at the given length, we seek to estimate general aspects of the landscape.*

For the asymmetric expansion (Fig. 6 Top), the 3_6 pseudoknot is dominant for 77 and 78nt and again around 89-98nt and 101-114nt. For other lengths, the 3_3 pseudoknot is dominant, namely over 95% of the landscape for 79-87nt (same for symmetric expansion with extra downstream nucleotides). In the alternative symmetric expansion (Fig. 6 Bottom), the probability of the 3_6 pseudoknot increases for 92-96nt, but drops for 98-104nt. After that, this conformation occupies almost the entire landscape for 106-142nt. A sudden switch to the dominant 3_3 conformation occurs at 144nt.

We choose the 87nt RNA with a probability of 99% 3_3 pseudoknot for further reactivity studies, which indeed yield a dominant 3_3. See Fig. S11. The dominant structure 4_21 is made of the 3_3 pseudoknot and the flanking Stem S_F, with a probability of 99.87%. Moreover, our partition algorithm[38] shows that 4_21 is a subgraph of 7_2192 (see Fig. S4), which corresponds to the 144nt 3_3-pseudoknot-containing structure in Fig. 4B. This indicates that our choice of 87nt preserves a natural structure adapted by the longer FSE while removing additional flanking nucleotides. The minor structure 4_12, a subgraph of 6_132 (144nt 3_6-pseudoknot-containing structure), is made of the 3_6 pseudoknot and a 5′ end hairpin.

We have also calculated landscapes using NUPACK and ShapeKnots without any SHAPE reactivities (Fig. S12). In those cases, many structures including 3_6, 3_3, 3_5, and 4_7 (the two-pseudoknot structure shown in Fig. 3 that includes the 3_3 and another pseudoknot at the 3′ end) emerge. For NUPACK, the 4_7 pseudoknot instead of 3_3 dominates 79-87nt, followed by a switch to a dominant 3_6 except at 98-104nt and 140-144nt for symmetric expansion. For ShapeKnots, only a small composition of 3_6 is seen; even for 77nt, we obtain a dominant 3_5 junction. Most of the landscape is occupied by 3_3.

What is clear from this landscape study is that the FSE conformation is length-dependent. Our confirmation of a 3_3 dominant landscape for 87nt by SHAPE reactivity data underscores the utility of the above analysis. The landscape analysis also emphasizes the flexibility of Stem 2 and suggests a possible switch between the two pseudoknot conformations. By design, the FSE of the virus may require conformational plasticity in the frameshifting task.

Comparison with other works

Other major and minor FSE conformations in the literature

To relate our three relevant, length-dependent structures for the FSE to recent structural works, we list major and minor FSE structures identified in Table S1. For SARS-CoV FSE, the 3_6 pseudoknot was taken as the consensus structure.[17,18,42] Since SARS-CoV-2 FSE only differs in one base, many studies presume that this FSE will continue to adopt the 3_6 conformation. Using various techniques and sequence lengths, 10 out of the 16 papers show a major 3_6 pseudoknot: iterative 2D prediction for 68nt;[43] 3D modeling and MD simulation for 68nt;[44] 2D prediction, NMR, DMS footprinting for 69nt;[45] 2D, 3D, and MD simulation for 77nt and 84nt;[20] small-angle X-ray scattering for 85nt;[19] DMS-MaPseq for 85nt;[22] homology model for 88nt;[46] Cryo-EM for 88nt[23] and 118nt;[21] deep sequencing for 1475nt.[47] All except the last two studies use short FSE lengths of 68-88nt. As we demonstrated here, short sequences, especially ≤77nt, tend to have a dominant 3_6 pseudoknot in the conformational landscape. The remaining 6 papers predict other major FSE conformations instead of the 3_6 pseudoknot: pseudoknot 3_8, or unknotted 2_1 and 2_2.

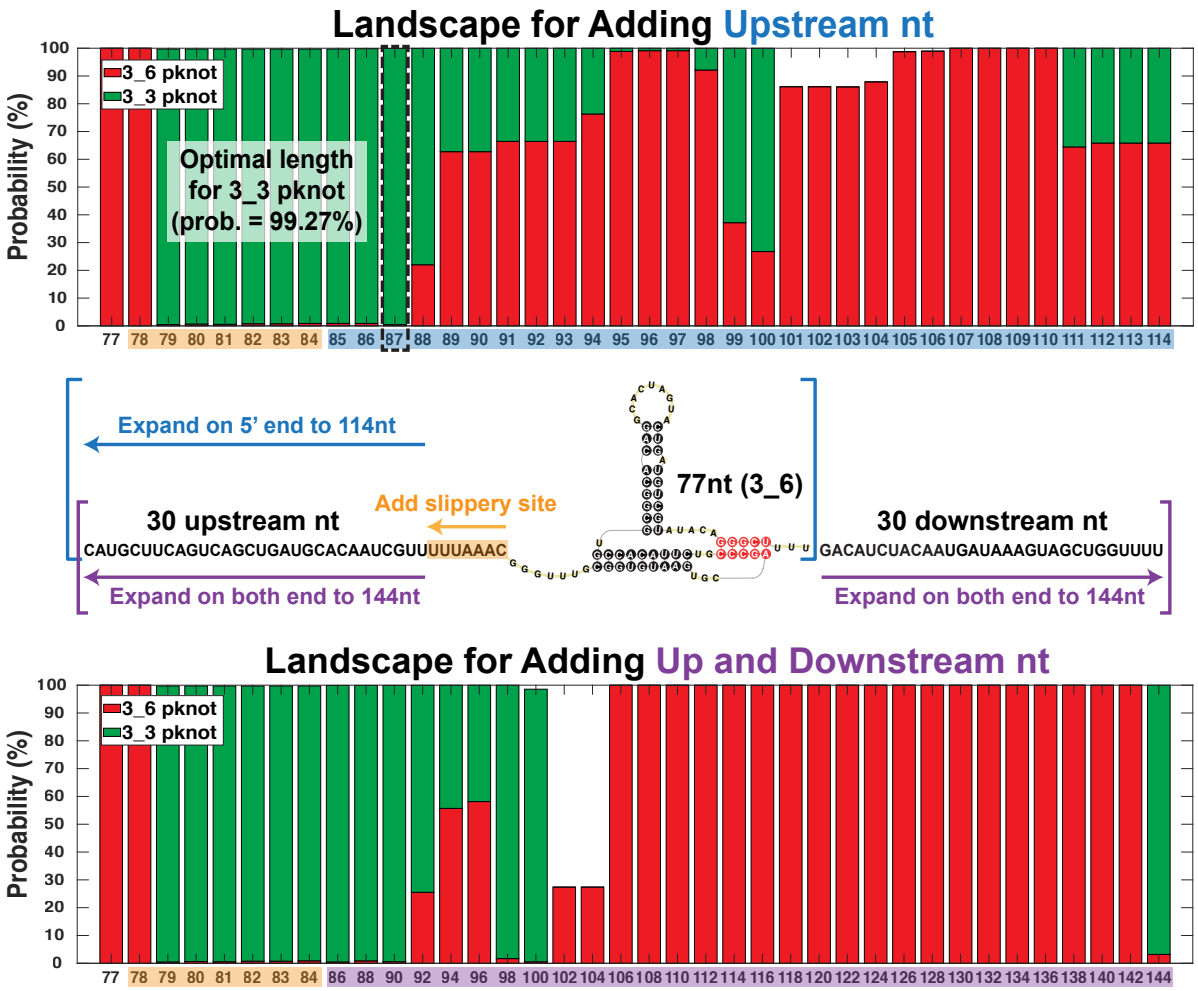


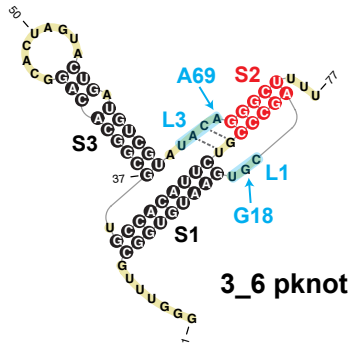
Figure 6: Conformational landscape of the frameshifting element for different sequence lengths predicted by ShapeKnots using reactivities from the 144nt construct. For each length, probabilities of all structures containing independently folded 3_6 or 3_3 pseudoknots are individually summed. The compositions are colored red (3_6) and green (3_3), respectively. (Top) Landscape for adding upstream nucleotides only to the 77nt FSE (asymmetric expansion). The optimal sequence length of 87nt for the 3_3 pseudoknot is in dashed black. (Bottom) Landscape for adding both upstream and downstream nucleotides to the 77nt FSE (symmetric approach). At 90nt (87nt + 3nt downstream) the landscape is almost all 3_3.

The 3_8 kissing hairpin for 126nt arises in an *in vivo* SHAPE experiment paper by Huston et al.,[4] where parameters for the 3_6 Stem 2 detected by ShapeKnots are hardwired constraints, and SuperFold[48] is applied to predict a consensus structure for the FSE. Their 3_8 pseudoknot (see Fig. S13) replaces the original Stem 3 by a different downstream stem, so that base pairs in the 3_6 Stem 2 involve loop regions of Stem 1 and this new stem.

The 2_1 unknotted conformation contains only Stems 1 and 3 (Fig. S13), so it is a substructure of all our conformers 3_6, 3_3, and 3_5 (Fig. S4). It is dominant in two studies using sequence length 81nt[49] and 123nt.[30] Both groups use a 2D prediction program which cannot predict pseudoknots, RNAfold.[50]

The 2_2 conformation is a two-stem structure with an internal loop, and it is derived by three groups all using 2D prediction programs with chemical reactivity data. ShapeKnots is used by both Manfredonia et al. with *in vivo* SHAPE (NAI reagent) data for 88nt,[8] and Zhang et al. with DMS and SHAPE (1M7 reagent) for 198nt.[23] Both groups predict a 2_2 structure that contains the 3_6 Stems 2 and 3 (Fig. S13). They expand only the 5′ end of the FSE, similar to our asymmetric approach for obtaining the conformational landscape (77-114nt). Our landscape does not produce 2_2, although the two-stem 2_1 is part of the landscape (white spaces) in Fig. 6. This can be explained by the longer (198nt) sequence used by Zhang et al.,[23] where Stem 1 is blocked by base pairing of the 77nt 5′ end and upstream nucleotides that go beyond our 114nt scope. The 3D structure built by Manfredonia et al.[8] from this 2_2 actually covers the 3_6 pseudoknot instead, so the report of the 2_2 could be an artifact of the 2D folding. Iserman et al. use SuperFold with SHAPE (5NIA reagent) data for 1000nt.[51] However, SuperFold does not predict pseudoknots, and

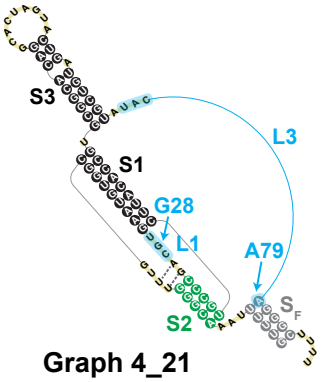
A



77nt		WT	G18A	G18C	Δ*L*1	ΔA69	Δ*L*3
Frameshifting efficiency of the WT level *		100%	30%	15%	0%	25%	10%
Prediction without SHAPE data	PKNOTS	3_6, S2 7bp	**4bp**	7bp	**6bp**	7bp	**4bp**
	NUPACK	3_6, S2 5bp	5bp	5bp	**4bp**	**7bp**	5bp
	IPknot	3_6, S2 5bp	5bp	**7bp**	**4bp**	**7bp**	5bp
	ProbKnot	3_5, S2 4bp	4bp	4bp	4bp	**7bp**	**3_3, S2 8bp**
ShapeKnots (From 77nt SHAPE Replicate 1)	3_6 pknot	97.67%	97.14%	99.6%	0.15%	100%	98.29%
	3_5 junction	2.33%	2.32%	0%	14.34%	0%	1.7%
	3_3 pknot	0%	0.54%	0.4%	85.51%	0%	0%

*Bhatt et al. (2020)

B



87nt		WT	G28A	G28C	Δ*L*1	ΔA79	Δ*L*3
Frameshifting efficiency of the WT level *		100%	30%	15%	0%	25%	10%
Prediction without SHAPE data	PKNOTS	4_12 (hairpin + 3_6 S2 7bp)	**4bp**	7bp	**6bp**	7bp	**4bp**
	NUPACK	4_12 (hairpin + 3_6 S2 5bp)	5bp	5bp	**4bp**	**7bp**	5bp
	IPknot	3_3, S2 7bp	7bp	**6bp**	**5bp**	7bp	7bp
	ProbKnot	3_3, S2 10bp	4_10 (hairpin + 3_5 S2 4bp)	4_10 (4bp)	4_10 (4bp)	**2_1**	**2_1**
ShapeKnots (From 87nt SHAPE)	4_12 (3_6 + hairpin)	0.13%	0.04%	5.11%	0%	100%	0.34%
	4_21 (3_3 + extra stem)	99.87%	99.96%	94.89%	100%	0%	99.66%

*Bhatt et al. (2020)

Figure 7: Effects of 5 mutations tested for frameshifting efficiency by Bhatt et al.[21] on 2D structure predictions of the 77nt and 87nt frameshifting element. (A) (Left) The 77nt FSE 3_6 pseudoknot with mutation regions labeled in blue. Two weak base pairs for Stem 2 are indicated using dotted lines. (Right) A table showing 2D prediction results for the wildtype and the mutants. The upper half is for 4 programs: 3_6, 3_5, and 3_3 predictions are in red, purple, and green, respectively, with their corresponding Stem 2 lengths (in bold if structure change). The lower half is for ShapeKnots, showing probabilities of 3_6, 3_5, and 3_3. (B) (Left) The 87nt FSE 2D structure by ShapeKnots, a 4_21 structure made of the 3_3 pseudoknot and a flanking Stem S_F, with mutation regions in blue. (Right) A table showing prediction results using 4 programs and ShapeKnots.

their 2_2 only contains Stem 3.

Many minor conformations have also been reported as summarized in SI, including our 3_3 pseudoknot and 3_5 junction captured by Zhang et al.,[23] who predict the 3_3 pseudoknot by ShapeKnots using DMS data for 88nt. However, DMS data can only inform about nucleotides A and C, which likely explains why 3_3 is only a minor conformer. Additionally, ShapeKnots is designed for SHAPE reactivities and not DMS. These researchers also obtain 3_5 using RNAstructure,[52] which again does not predict pseudoknots.

Trinity et al.[43] obtain a minor 3_8 and a 3_3 pseudoknot (Fig. S13), but the corresponding 2D structures are different from those we have described. Recall that each graph topology corresponds to multiple 2D structures. In their 3_8 RNA,[43] base pairs in Stems 2 and 3 are the same as in our 3_6 in Fig. 1, but the original Stem 1 is replaced by a pseudoknot that binds the 5′ end and the Stem 3 loop. In their 3_3,[43] Stem 3 loop and the 3′ end base pair to form the pseudoknot, instead of the 5′ end of the 77nt FSE and Stem 1 loop. The 2_2 unknotted RNA conformation appears again but as a minor conformer by Lan et al.,[22] who derive it using ShapeKnots with DMS data for 283nt.

FSE structure and frameshifting efficiency

While clearly the conformational landscape of the FSE is length dependent and fragile to mutations, the relation between structure and frameshifting efficiency is not well understood. In the Cryo-EM study by Bhatt et al.,[21] the researchers show that mutations of a single residue G13486 in Loop 1 (Fig. 1) reduce frameshifting efficiency, and that deletion of the entire Loop 1 inhibits frameshifting entirely; similarly, removing a single residue A13537 in Loop 3 or the entire Loop 3 reduces frameshifting dramatically (Fig. 7). They suggest that such changes in frameshifting efficiency are caused by altered interactions with the ribosome.

To investigate whether these mutations might also affect the FSE structure and possibly the frameshifting process through a structural change, we predict for each

mutation in Fig. 7 the 2D structures of the resulting RNAs for 77nt and 87nt FSEs using four prediction programs (PKNOTS, NUPACK, IPknot, and ProbKnot). We also consider predictions with reactivity data for the relevant residues in the original 77nt and 87nt FSE constructs. We use red/green/purple consistent with Fig. 1 to highlight resulting 3_6/3_3/3_5 conformations.

Without reactivity data, while the three programs that predict the 3_6 pseudoknot for the wildtype 77nt FSE continue to predict 3_6 for the mutants, Stem 2's length is altered. Meanwhile, ProbKnot predicts the 3_5 junction with a 4bp Stem 2 for the wildtype FSE. This 3_5 Stem 2 is lengthened by deletion of A13537 in Loop 3, but destroyed by deletion of the entire Loop 3, where the alternative 3_3 Stem 2 forms. For 87nt systems, the 3_3 pseudoknot emerges as dominant structure for the IPknot program for both wildtype and all mutants, consistent with our predictions and SHAPE experiments (Fig. 6 and S11), while only the wildtype for ProbKnot; the structures predicted by ProbKnot for the mutants yield both the 3-way junction and a simple two-stem structure. With SHAPE data, we see that the 3_3 pseudoknot dominates over 3_6 when Loop 1 is deleted, and that all three conformations again play a role in the energy landscape.

Taken together, with the analysis above, many alternative conformations appear in the literature for the FSE. Even using similar methods such as chemical structure probing can lead to different conformations for different sequence lengths; this again underscores how sequence length matters in predicting this FSE structure. From both our mutant analysis and the 2D structure predictions for the Bhatt et al. mutants,[21] we conclude that the FSE structure is highly sensitive to mutations. Altering only a few nucleotides can transform the FSE conformation to an alternative structure or decrease the length of a stem significantly, possibly reducing the frameshifting efficiency.

Conclusion and discussion: mechanistic implications

Using a combination of graph-based modeling, 2D structure prediction programs, and chemical structure probing data, we have described three alternative structures for the SARS-CoV-2 frameshifting element (Fig. 1). Besides the 3-stem H-type pseudoknot long assumed to be the dominant structure in the literature (3_6 dual graph),[19,24] we highlight another 3-stem pseudoknot, HL-type (3_3 dual graph) that becomes dominant when the slippery site's 7 residues or more, are considered. In addition, an unknotted 3-way junction RNA (3_5 dual graph) also emerges as a minor player in the FSE conformational landscape. Using minimal mutations predicted by our genetic algorithm RAG-IF, we can strengthen the prevalence of 3_6, 3_3, and 3_5 in the 77nt construct using six, three, and two mutations, respectively, highlighting the fragility of the sequence/structure relationship for the FSE. The SHAPE reactivity results summarized in Table 1 for all wildtype and mutant replicates confirm our predictions experimentally.

The two main pseudoknot structures are differently intertwined by hydrogen bonding: Stem 1's loop base pairs with the 3′ end of the FSE to form Stem 2 in the 3_6 pseudoknot, while this loop binds the 5′ end of the FSE in 3_3 (Fig. 1). Both are bulky structures because of the perpendicular orientation of Stem 3 with respect to the Stem 1/Stem 2 plane and have different inherent motions (see three-dimensional views in Fig. 8, Methods, and more details in a separate molecular dynamics paper[53]). Our estimated conformational landscape as a function of sequence length (Fig. 6) further highlights the plasticity of the FSE, and the likelihood that exogenous factors, such as small molecules, will alter the conformational landscape.

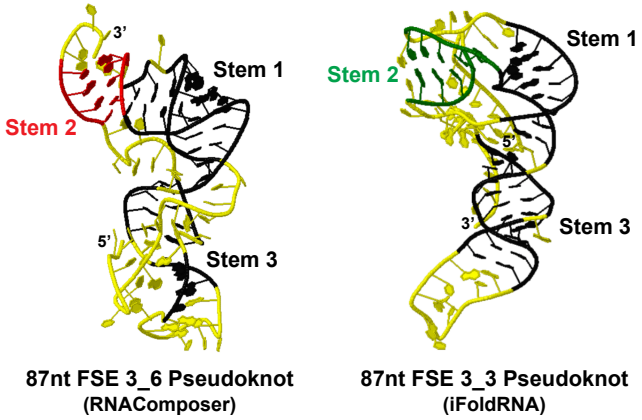


Figure 8: Three-dimensional models of the 87nt 3_6 and 3_3 pseudoknot. Initial 3D structures are predicted by RNAComposer,[54] Vfold3D,[55] SimRNA,[56] and iFoldRNA,[57] and subjected to 1-1.5μs MD using Gromacs.[58] The last 500ns are used for clustering analysis, and the most populated cluster center by RNAComposer/iFoldRNA is shown here for each system.

Our multiple sequence alignment of coronaviruses (Fig. 2) similarly underscores the variability of Stem 2 among this virus family. Sequence similarity of the FSE segment (60-99%) is higher than the overall genome similarity (54-79%) among these family members, especially for SARS, OC43, and MHV, emphasizing the high conservation of the FSE gene element. The flexibility of Stem 2 suggests that this region of the FSE may be involved in a switch between the alternative conformations and possibly associations with the ribosome during transcription and protein translation (see below).

While our predictions using 2D folding programs and chemical reactivity data do not identify other conformations besides the two pseudoknots and the 3-way junction, other possibilities cannot be excluded due to limitations in both the folding programs and reactivity experiments. Results for other works suggest other

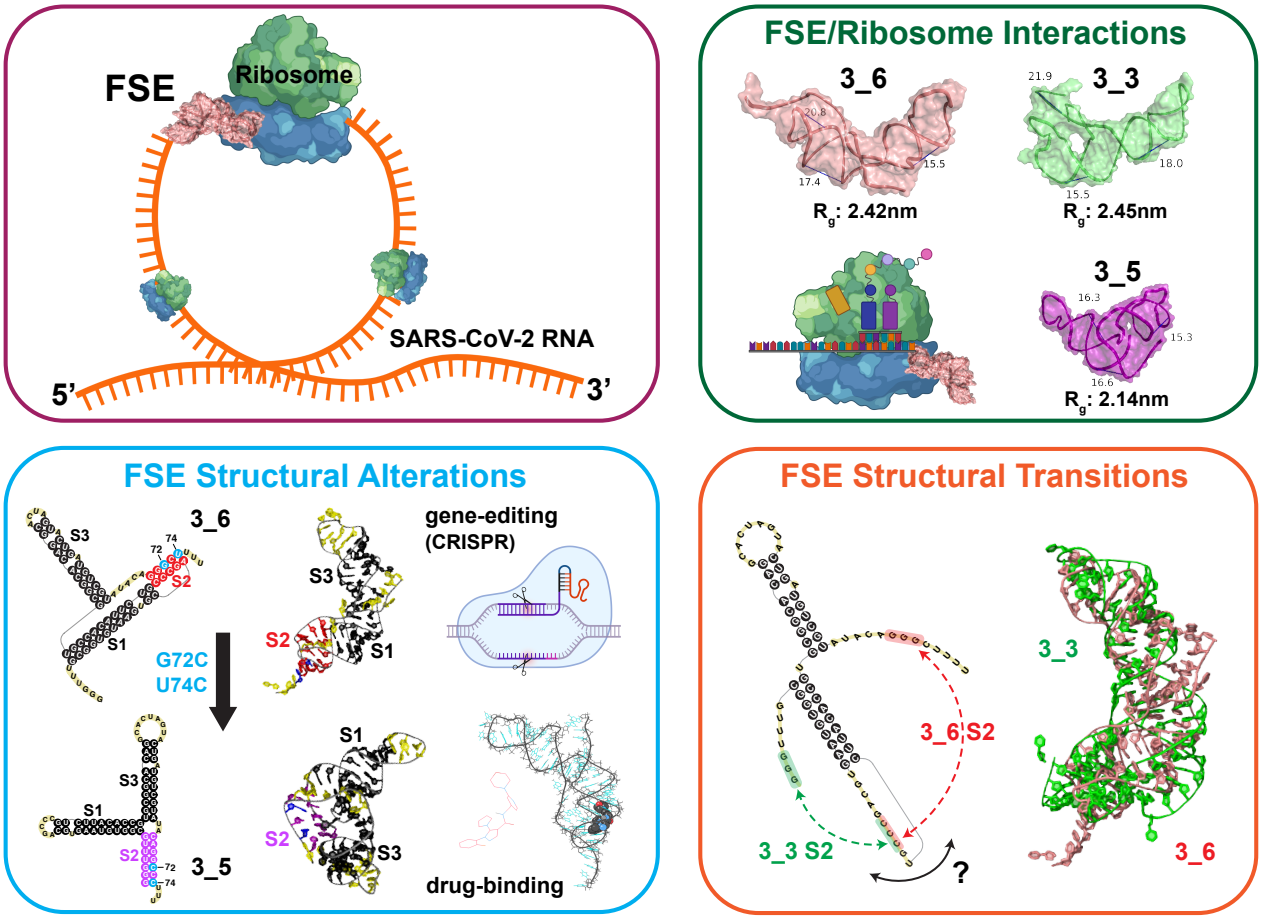


Figure 9: Three avenues for frameshifting interference.**

possible conformations, including two 3_8 pseudoknots, a different 3_3 pseudoknot (with Stem 2 formed by Stem 3 loop and 3′ end), a two-stem 2_1, and a two-stem-with-internal-loop 2_2 (Table S1 and Fig. S13).

The length-dependent and context-specific conformations for the FSE emphasized in our work could be exploited biologically by the insidious virus in mechanisms of interactions with the ribosome. The bulky pseudoknot may block and encourage ribosome pausing as the ribosome rolls over the FSE region, and any structural transitions may similarly interfere with this template-directed process.

In the 2.3–7Å-resolution Cryo-EM study by Bhatt et al.,[21] the researchers observe the 3_6 pseudoknot wedged between the head and body of the small ribosomal subunit. Specifically, the residues before Stem 1 are found to be located at the mRNA entry tunnel so that, as translation proceeds, the GC-rich Stem 1 residues are unwound upon entering the mRNA tunnel. Stem 2 of the pseudoknot is found to be shorter by two base pairs than that predicted by computational 2D structure methods. To confirm these inferred structural interactions, the researchers mutated single residues in Loops 1 and 3, and showed subsequent reduction in the frameshifting efficiency. Our corresponding 2D structure predictions for these mutants reveal that the FSE structure is highly sensitive to mutations (prior section). The ribosome anchoring is likely to reduce this conformational variability in the realistic context, but the bulkiness of the pseudoknot may be part of the structural signalling as the ribosome unwinds the FSE. Mutations and drug interference with the FSE can thus play a crucial role in affecting frameshifting efficiency. We can thus envision at least three avenues for such interference (Fig. 9).

FSE Mutations: For the isolated FSE, mutations of residues by gene editing or drug binding that alter the FSE structure can interfere with frameshifting. This was proposed in our prior work, where minimal mutations of 1–2 residues were predicted.[20] Drug binding screening *in silico* to these identified mutated residues are already identifying candidate compounds (In Preparation[59]).

FSE/Ribosome: A second avenue of intervention can be envisioned by influencing the FSE/ribosome interactions. Loop 1 or Stem 1 could be good targets here for mutations or drug binding to interfere with the biomolecular recognition process and protein translation occurring in the mRNA entry tunnel.[21]

FSE Transitions: Finally, influencing structural transitions of the FSE could be a possible route for intervention, but more information is required about the natural dynamics of this intriguing system. With the many structures described in the literature, complex struc-

tural rearrangements could be involved that may be context specific. Insights into this heterogeneous landscape could come from atomic-level molecular dynamics simulations. Already, our microsecond simulations suggest different inherent motions and threading orientations for the two FSE pseudoknot systems (In Preparation[53]). Importantly for small molecule targeting, stabilization of any of the three conformations is likely to impact frameshifting.

As more high resolution structural complexes are reported, it may be possible to hone this picture. Computational studies will clearly be an important part of piecing the clues, as already demonstrated for many aspects of the Covid-19 disease. A combination of coarse-grained modeling as used here for RNA representations with RAG graphs and efficient design of minimal structure-altering mutations, are particularly effective when combined with atomic-level views.

Despite recent suggestions that viral mutations in the spike-protein encoding region may be associated with higher infectivity of recent variants, mutations in the FSE in these variants (Fig. S2) appear random and infrequent, reinforcing this region's high evolutionary conservation and importance to maintaining viral fitness. Of course, should significant variants arise, we can easily extend the current modeling framework to those. Clearly, the sequence-length and context dependence folding of the FSE and its conformational variability make modeling and experiments infinitely more complicated, but these variations may be a part of the complex machinery that coronaviruses have developed to infect and replicate rapidly and efficiently.

Even as we enter a brighter period of the current pandemic with arrival of multiple highly effective vaccines against Covid-19, the threat of further variants, coronavirus waves, and other viruses cannot be overstated. With increased global travel, human invasion of natural forests, and domestication and consumption of wild animal species, more opportunities arise for the jumping of viruses from their natural reservoirs in the animal kingdom to human hosts. A better understanding of the complex structure/function relationship will be critical in this fight against future virus pandemics. With continued collaborations, we will likely be far more prepared.

Methods

SARS-CoV-2 RNA sequences

We use the official SARS-CoV-2 RNA reference sequence provided by GISAID[27] (Accession ID: EPI_ISL_402124), 29891nt. The 84nt FSE occupies residues 13462–13545, and the spike gene region is residues 21564–25384. Other viral RNA sequences are aligned to the reference by GISAID using *mafft*.[60] Specific variants such as the British variant (B.1.1.7 or VUI-202012/01), the South Africa variant (B.1.351 or 501Y.V2), the Brazil variant (P.1), or the New York City variant (B.1.526) were downloaded as fasta files using its search engine.

Coronavirus FSE multiple sequence alignment

There are four steps in our coronavirus FSE MSA:

1. *Coronavirus selection*: We use the 9 coronaviruses identified in a SARS −1 programmed ribosomal frameshifting (PRF) homology search[42] with the RECODE database.[61] We also check RECODE for the −1 PRF signal in coronaviruses for updates. We add SARS and MERS along with the 9 coronaviruses. Sequences for the 11 coronaviruses are downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) with accession numbers: SARS (NC_004718), OC43 (KU131570), MHV (NC_001846), BCoV (NC_003045), HKU1 (NC_006577), MERS (KT225476), AIBV (NC_001451), NL63 (NC_005831), FIP (NC_002306), 229E (NC_002645), PEDV (NC_003436).

2. *FSE region identification*: For each coronavirus, we use local pairwise alignment (gap opening penalty −2, gap extension penalty −1) to identify homologous regions to the 84nt SARS-CoV-2 FSE regions.

3. *Homology calculation*: Sequence similarities with SARS-CoV-2 for both the whole genome and the 84nt FSE segment are calculated using BLAST's global alignment webserver with default parameters.[62]

4. *MSA*: We apply the multiple sequence alignment software ClustalW2[28] to align the 84nt coronavirus FSEs, and display them using Jalview.[63]

Secondary structure prediction programs

Six 2D structure prediction programs that can handle pseudoknots are used: PKNOTS,[33] NUPACK,[34] IPknot,[35] ProbKnot,[36] vsfold5,[37] and ShapeKnots.[40] Only ShapeKnots can incorporate SHAPE reactivities into the prediction of 2D structures. Except for vsfold5, which works as a webserver, we install the programs locally. Default parameters are used for PKNOTS, NUPACK *mfe*, ProbKnot, and vsfold5. For IPknot, the parameters are set to level 2 prediction, CONTRAfold scoring model, refinement 1, and base pair weights 2 for level 1 and 16 for level 2. For ShapeKnots, we provide experimental SHAPE reactivity data as input, and calculate all suboptimal structures.

Folding predictions for 4 pseudoknot-containing RNAs

To examine how sequence lengths affect 2D structure prediction programs, we selected four experimentally solved RNAs with pseudoknots from the Protein Data Bank (PDB) https://www.rcsb.org. The RNAs are:

lysine riboswitch (PDB ID: 3DIO), lariat-capping ribozyme (6GYV), glmS ribozyme (2GCV), and T-Box riboregulator (6UFG). The 3D structures are extracted from respective PDB files, and corresponding 2D structures are determined using the RNApdbee webserver.[64]

Pseudoknot substructures in these RNAs are identified and taken to define "short" systems for 2D structure prediction. An equal number of nucleotides are added to both ends of the lysine riboswitch and lariat-capping ribozyme pseudoknots to reach 120nt. Only the 3′ ends of the glmS ribozyme and T-Box riboregulator pseudoknot are expanded to 120nt, because no upstream nucleotides are relevant. The five programs PKNOTS, NUPACK, IPknot, ProbKnot, and vsfold5 are used to predict 2D structures for three sizes ("short", 120nt, and all). The predictions are compared to the 2D structures extracted from the experimental structures to assess the number of residues that are correctly base paired. The accuracies are calculated as the percentage of correctly base-paired residues in the total number of residues.

SHAPE, PairMap

Synthesis and purification of *in vitro* transcribed RNA:

Various constructs of SARS-CoV-2 FSE (with and without slippery site) were synthesized from Integrated DNA Technologies (IDT). Each construct is flanked by structural RNA adapters[65] and a T7 promoter region at the 5′ end. These DNA constructs were used as a template to *in vitro* transcribed RNA using T7 high yield RNA kit (New England Biolabs). The synthesized RNA was DNase treated (TURBODNase), purified using Purelink RNA mini kit (Invitrogen) and quantified with nanodrop.

Modification of *in vitro* transcribed RNA

Samples of 6 μg of *in vitro* transcribed RNA was denatured at 65°C for 5 minutes and snap-cooled in ice. After the addition of folding buffer (100 mM KCl, 10 mM MgCl2, 100 mM Bicine, pH 8.3), RNA was incubated at 37°C for 10 minutes. The folded RNA was treated with either 10μl of Dimethyl sulfate (DMS, 1:10 ethanol diluted) or with 10ul of 5-Nitro Isatoic Anhydride (5NIA, 25mM final concentration). Subsequently, for negative controls (unmodified RNA) equivalent amount of ethanol and Dimethyl Sulfoxide (DMSO) was added to the folded RNA. The complete reaction mixture was for further incubated for 5 minutes at 37°C to allow complete modifications of the unpaired RNA nucleotides. DMS treated reaction was quenched using 100μl of 20% beta-mercaptoethanol (β-ME). Both the modified and unmodified RNAs were purified using the PurelinkRNA mini kit and quantified with nanodrop.

cDNA synthesis, library construction, sequencing, and data processing

Purified RNA from above was reverse transcribed using Gene-specific reverse primer (Table 2) directed against the 3′ RNA adapter sequence and SuperScript II reverse transcriptase under error prone conditions as previously described.[48] The resultant cDNA was purified using G50 column (GE healthcare) and subjected to second strand synthesis (NEBNext Second Strand Synthesis Module). For library generation, we designed primers, specific to the 5′ and 3′ RNA adapter sequence (Table 2) and PCR amplified the whole cDNA using the NEB Q5 HotStart polymerase (NEB). Secondary PCR was performed to introduce TrueSeq barcodes.[48] All samples were purified using the Ampure XP (Beckman Coulter) beads and Quantification of the libraries was done using Qubit dsDNA HS Assay kit (ThermoFisher). Final libraries were run on Agilent Bioanalyzer for quality check. These TrueSeq libraries were then sequenced as necessary for their desired length, primarily as paired end 2×151 read multiplex runs on MiSeq platform (Illumina). We used the ShapeMapper2 algorithm[66] to determine the mutation frequency in both chemically modified (5NIA and DMS treated) and control (DMSO and ethanol treated) RNA samples and to calculate chemical reactivity for each RNA nucleotide using the following equation:

$$R = mutr_m - mutr_u,$$

where R is the chemical reactivity, $mutr_m$ is the mutation rate calculated for chemically modified RNA and $mutr_u$ is the mutation rate calculated for untreated control RNA samples.

Shapemapper2 was also used to calculate the parse mutations from DMS-MaP sequencing data.[66] The resulting parsed mutation files were used in the Pair-MaP pipeline which uses PairMapper and RingMapper to compute and identify correlated mutations in the DMS-MaP sequencing dataset.[41] The correlated mutational outputs were plotted with arcPlot.[41]

SHAPE and DMS reactivities calculated for all the wildtype and mutant constructs are available in the supplementary file *SARS-CoV-2-FSE_SNRNASM.xlsx.*

RAG-IF and mutants design

RAG-IF for dual graphs

RAG-IF is an RNA-As-Graphs based inverse folding program that uses genetic algorithm to mutate an RNA sequence so that it folds onto a different target structure (graph) by minimal mutations. It was originally designed and fully automated for tree graphs,[26] and modified for dual graphs with manual intervention to select the mutation regions.[20] Two prediction programs that can handle pseudoknots are used to determine folding success. Our default options are IPknot and NUPACK (programs A and B, respectively). However, for 3_6

Table 2: Primers used for the SuperScript II error prone Reverse Transcriptase PCR and library generation

Primer	Sequence
3′ Cassette-RT	GAACCGGACCGAAGCCCG
5′ Cassette-Fwd	CCCTACACGACGCTCTTCCGATCTNNNNNGCCTTCGGGCCAA
3′ Cassette-Rev	GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNGAACC GGACCGAAGCCCG

pseudoknot-strengthening mutant, only ProbKnot predicts a graph (3_5) that is not 3_6 for the wildtype 77nt. Hence, we substitute default IPknot by ProbKnot in the design of 3_6 PSM. RAG-IF has three steps:

1. *Mutation regions and target structure*: We identify the smallest mutation region for breaking or forming stems to fold onto the target graph and design a target 2D structure for the target graph.

2. *Genetic algorithm*: We create an initial population of N sequences by randomly assigning nucleotide identities to the mutation regions. Each individual sequence then receives a *fitness* score, which is the number of residues predicted by *Program A* to have the same 2D structure as the target folding, as calculated by the Hamming distance. This population is then subject to k iterations of random mutation, crossover, and selection, and those with high fitness are retained as candidates. The algorithm stops once we have enough candidates or the execution time is too long. These candidate sequences are further screened by *Program A* and *B*, and only those that fold onto the target graph by both programs are retained.[26]

3. *Optimization*: For each sequence survived above, we remove unnecessary mutations, i.e., the sequence still folds onto the target graph by both prediction programs without these mutations. The remaining mutations are considered minimal.[26]

3_6 pseudoknot-strengthening mutant

We apply RAG-IF to the 77nt FSE sequence to predict a 3_6 pseudoknot-strengthening mutant, as illustrated in Fig. S7. ProbKnot predicts a 3_5 junction for the wildtype 77nt (Fig. S7A). The mutation regions are: residues 1-3 to avoid alternative 3_3 pseudoknot Stem 2, and residues 4 and 67-69 to break the 3_5 Stem 2. For the target 2D structure, we use the 3_6 structure by ShapeKnots (Fig. 4A) with shortened Stem 1. For the genetic algorithm, we create a population of 500 sequences ($N = 500$), and $k = 500$ iterations. The program is terminated when at least 500 candidates are produced or the execution time exceeds 12 hours. RAG-IF generates 75 unique sequences with 2-6 mutations. The results are listed in Fig. S7A with illustrative mutations.

To dominate the landscape with the 3_6 conformation (rather than obtain just minimal mutations), we also examine the strength of Stem 2. By screening the 75 mutated sequences by 4 prediction programs (PKNOTS, NUPACK, IPknot, and ProbKnot), we identify the quadruple-mutant [G3U, U4A, C68A, A69C] that all four 2D structure-prediction programs fold onto 3_6 (Fig. S7A). Stem 2 has 7 base pairs using three programs and 5 for ProbKnot. To further strengthen Stem 2, we also mutate residues 18 and 19 in Stem 1 to A, so that they base pair with the UU in the 3′ end. With 6 mutations [G3U, U4A, G18A, C19A, C68A, A69C], all programs predict 9 base pairs for Stem 2.

We test the above 6 mutations on the 144nt construct. PKNOTS, NUPACK, and IPknot give similar structures as the suboptimal 6_132 structure by ShapeKnots (Fig. 4B): two hairpins in the 5′ end, followed by the 3_6 pseudoknot, and finally a hairpin (highlighted green in Fig. S7B) in the 3′ end. However, ProbKnot predicts a pseudoknot-free structure with only Stems 1 and 3 of 3_6. The 3′ strand of Stem 2 (pink) forms a different stem with the 3′ end. To break this stem and restore 3_6 Stem 2, we add mutation C137A to destroy the middle GC base pair, without altering Stem 2 (pink) and the 3′ end hairpin (green). As expected, the 7-mutant FSE [G40U, U41A, G55A, C56A, C105A, A106C, C137A] yields similar structures containing a 3_6 pseudoknot with 9 base pairs for Stem 2 by all four programs.

3_3 pseudoknot-strengthening mutant

To stabilize the 3_3 pseudoknot, we apply RAG-IF to the 77nt FSE (Fig. S9) with mutation regions defined by residues 4-6 to form a strengthened 3_3 Stem 2, and residues 70-73 to break the 3_6 Stem 2. We use the same parameters for the genetic algorithm as above. We obtain 20 unique sequences with 1-4 mutations, listed in Fig. S9. After 2D prediction program screening, we consider the triple-mutant [U4C, G71A, G72U] to be the strongest, with 5-7 base pairs for Stem 2 of 3_3.

3D models

The 3D structures of the 87nt FSE were predicted using RNAComposer,[54] Vfold3D,[55] SimRNA,[56] and iFoldRNA.[57] One structure from each with the correct graph topology was used for MD simulations.

MD simulations were performed using Gromacs 2020.4,[58] with the Amber OL3 forcefield.[67] The systems were solvated with TIP3P water molecules in the cubic box whose boundaries extended at least 10 Å from any RNA atom.[68] After charge neutralization with randomly placed sodium ions, additional Na^+ and

Cl^- ions were added for 0.1M bulk concentration. The systems were energy minimized via steepest descent and equilibrated with position restraints on the RNA. Simulations were run with a timestep of 2 fs and a SHAKE-like LINCS algorithm[69] with constraints on all bonds. The Particle Mesh Ewald method[70] was used to treat long-range electrostatics. The equilibration was performed for 100 ps in the NVT ensemble (300 K) and then 100 ps in NPT ensemble (300 K and 1 bar). The RNA and ionic solvent were independently coupled to external heat baths with a relaxation time of 0.1 ps. Production runs were performed for at least 1 μs under NPT, based on when the RMSD stabilized.

Cluster analysis was conducted via Gromos using conformations every 200ps within the last 500ns in each simulation using RNA non-H backbone atoms. With a cutoff of 3 Å, the largest cluster occupies 82.3% and 22.4% for 3_6 and 3_3, respectively. The 5′ end in 3_6 is threaded through the ring formed by the 3 stems and extends along the strand of Stem 3 with residues stacked with Stem 3 residues. The 5′ end in 3_3 is not threaded and instead forms a new Stem 2 and pairing with 3′ end. Details of the new MD simulations are described separately in.[53]

Associated Content

Supporting Information

SARS-CoV-2 mutation maps; Length effects on 2D predictions; Dual graph partition; Replicate 2 SHAPE analysis for 77nt and 144nt FSE; RAG-IF mutant designs; SHAPE reactivity differences; SHAPE replicate alignments; SHAPE analysis for 87nt FSE; Additional FSE conformational landscape; Other FSE conformations reported in the literature

SARS-CoV-2-FSE_SNRNASM.xlsx

SHAPE and DMS reactivities calculated for all the wild-type and mutant constructs

Author Information

Corresponding Authors

Tamar Schlick - *Department of Chemistry, New York University, New York, New York 10003, United States; Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States; NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, P.R. China; Email:* schlick@nyu.edu

Alain Laederach - *Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States; Email:* alain@unc.edu

Authors

Qiyao Zhu - *Courant Institute of Mathematical Sciences, New York University, New York, New York 10012, United States*

Abhishek Dey - *Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, United States*

Swati Jain - *Department of Chemistry, New York University, New York, New York 10003, United States*

Shuting Yan - *Department of Chemistry, New York University, New York, New York 10003, United States*

Notes

The authors declare no competing financial interest.

Acknowledgments

We gratefully acknowledge funding from the National Science Foundation RAPID Award 2030377 from the Divisions of Mathematical Science and of Chemistry, National Institutes of Health R35GM122562 Award from the National Institute of General Medical Sciences, and Philip-Morris International to T. Schlick, and National Institutes of Health grants R01 GM101237 and R01 HL111527 to A. Laederach.

References

(1) Yao, H.; Song, Y.; Chen, Y.; Wu, N.; Xu, J., et al. Molecular Architecture of the SARS-CoV-2 Virus. *Cell* **2020**, *183*, 730–738.e13.

(2) Turoňová, B.; Sikora, M.; Schürmann, C.; Hagen, W. J. H.; Welsch, S., et al. In situ structural analysis of SARS-CoV-2 spike reveals flexibility mediated by three hinges. *Science* **2020**, *370*, 203–208.

(3) Casalino, L.; Gaieb, Z.; Goldsmith, J. A.; Hjorth, C. K.; Dommer, A. C., et al. Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein. *ACS Cent. Sci.* **2020**, *6*, 1722–1734.

(4) Huston, N. C.; Wan, H.; Strine, M. S.; de Cesaris Araujo Tavares, R.; Wilen, C. B.; Pyle, A. M. Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **2021**, *81*, 584–598.e5.

(5) Zhang, Y.; Hong, S.; Ruangprasert, A.; Skiniotis, G.; Dunham, C. M. Alternative Mode of E-Site tRNA Binding in the Presence of a Downstream mRNA Stem Loop at the Entrance Channel. *Structure* **2018**, *26*, 437–445.

(6) Schubert, K.; Karousis, E. D.; Jomaa, A.; Scaiola, A.; Echeverria, B., et al. SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nat. Struct. Biol.* **2020**, *27*, 959–966.

(7) Thoms, M.; Buschauer, R.; Ameismeier, M.; Koepke, L.; Denk, T.; Hirschenberger, M., et al. Structural basis for translational shutdown and immune evasion by the Nsp1 protein of SARS-CoV-2. *Science* **2020**, *369*, 1249–1255.

(8) Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T. K., et al. Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **2020**, *48*, 12436–12452.

(9) Smith, M.; Smith, J. Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface. 2020; ChemRxiv, doi: 10.26434/chemrxiv.11871402.v3, preprint posted March 2020.

(10) Nguyen, T. M.; Zhang, Y.; Pandolfi, P. P. Virus against virus: a potential treatment for 2019-nCov (SARS-CoV-2) and other RNA viruses. *Cell Res.* **2020**, *30*, 189–190.

(11) Miao, Z.; Adamiak, R. W.; Antczak, M.; Boniecki, M. J.; Bujnicki, J. M., et al. RNA-Puzzles Round IV: 3D structure predictions of four ribozymes and two aptamers. *RNA* **2020**, *26*, 982–995.

(12) Sun, L.-Z.; Zhang, D.; Chen, S.-J. Theory and Modeling of RNA Structure and Interactions with Metal Ions and Small Molecules. *Annu. Rev. Biophys.* **2017**, *46*, 227–246.

(13) Deigan, K. E.; Li, T. W.; Mathews, D. H.; Weeks, K. M.; Tinoco, I. Accurate SHAPE-Directed RNA Structure Determination. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 97–102.

(14) Smola, M. J.; Weeks, K. M. In-cell RNA structure probing with SHAPE-MaP. *Nat. Protoc.* **2018**, *13*, 1181–1195.

(15) Brierley, I. Ribosomal frameshifting on viral RNAs. *J. Gen. Virol.* **1995**, *76*, 1885–1892.

(16) Atkins, J. F.; Loughran, G.; Bhatt, P. R.; Firth, A. E.; Baranov, P. V. Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **2016**, *44*, 7007–7078.

(17) Park, S.-J.; Kim, Y.-G.; Park, H.-J. Identification of RNA Pseudoknot-Binding Ligand That Inhibits the -1 Ribosomal Frameshifting of SARS-Coronavirus by Structure-Based Virtual Screening. *J. Amer. Chem. Soc.* **2011**, *133*, 10094–10100.

(18) Ritchie, D. B.; Soong, J.; Sikkema, W. K. A.; Woodside, M. T. Anti-frameshifting Ligand Reduces the Conformational Plasticity of the SARS Virus Pseudoknot. *J. Amer. Chem. Soc.* **2014**, *136*, 2196–2199.

(19) Kelly, J. A.; Olson, A. N.; Neupane, K.; Munshi, S.; Emeterio, J. S.; Pollack, L.; Woodside, M. T.; Dinman, J. D. Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *J. Biol. Chem.* **2020**, *295*, 10741–10748.

(20) Schlick, T.; Zhu, Q.; Jain, S.; Yan, S. Structure-Altering Mutations of the SARS-CoV-2 Frameshifting RNA Element. *Biophys. J.* **2021**, *120*, 1040–1053.

(21) Bhatt, P. R.; Scaiola, A.; Loughran, G.; Leibundgut, M.; Kratzel, A., et al. Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. **2020**, bioRxiv, doi: 10.1101/2020.10.26.355099, preprint posted October 2020.

(22) Lan, T. C. T.; Allan, M. F.; Malsick, L. E.; Khandwala, S.; Nyeo, S. S. Y., et al. Insights into the secondary structural ensembles of the full SARS-CoV-2 RNA genome in infected cells. **2021**, bioRxiv, doi: 10.1101/2020.06.29.178343, preprint posted February 2021.

(23) Zhang, K.; Zheludev, I. N.; Hagey, R. J.; Wu, M. T.-P.; Haslecker, R., et al. Cryo-electron Microscopy and Exploratory Antisense Targeting of the 28-kDa Frameshift Stimulation Element from the SARS-CoV-2 RNA Genome. **2020**, bioRxiv, doi: 10.1101/2020.07.18.209270, preprint posted July 2020.

(24) Rangan, R.; Zheludev, I. N.; Hagey, R. J.; Pham, E. A.; Wayment-Steele, H. K.; Glenn, J. S.; Das, R. RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *RNA* **2020**, *26*, 937–959.

(25) Chen, S. J. Graph, pseudoknot, and SARS-CoV-2 genomic RNA: A biophysical synthesis. *Biophys. J.* **2021**, *120*, 980–982.

(26) Jain, S.; Tao, Y.; Schlick, T. Inverse Folding with RNA-As-Graphs Produces a Large Pool of Candidate Sequences with Target Topologies. *J. Struct. Biol.* **2020**, *209*, 107438.

(27) Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **2017**, *1*, 33–46.

(28) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A., et al. Clustal W and Clustal X version 2.0. *Bioinform.* **2007**, *23*, 2947–8.

(29) Tavares, R. d. C. A.; Mahadeshwar, G.; Wan, H.; Huston, N. C.; Pyle, A. M. The global and local distribution of RNA structure throughout the SARS-CoV-2 genome. *J. Virol.* **2020**, *95*, e02190–20.

(30) Andrews, R. J.; Peterson, J. M.; Haniff, H. S.; Chen, J.; Williams, C.; Grefe, M.; Disney, M. D.; Moss, W. N. An in silico map of the SARS-CoV-2 RNA Structurome. **2020**, bioRxiv, doi: 10.1101/2020.04.17.045161, preprint posted April 2020.

(31) Lange, S. J.; Maticzka, D.; Möhl, M.; Gagnon, J. N.; Brown, C. M.; Backofen, R. Global or local? Predicting secondary structure and accessibility in mRNAs. *Nucleic Acids Res.* **2012**, *40*, 5215–5226.

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

(33) Rivas, E.; Eddy, S. R. A dynamic programming algorithm for RNA structure prediction including pseudoknots. *J. Mol. Biol.* **1999**, *285*, 2053–2068.

(34) Dirks, R. M.; Pierce, N. A. A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem* **2003**, *24*, 1664–1677.

(35) Sato, K.; Kato, Y.; Hamada, M.; Akutsu, T.; Asai, K. IPknot: fast and accurate prediction of RNA secondary structures with pseudoknots using integer programming. *Bioinform.* **2011**, *27*, i85–i93.

(36) Bellaousov, S.; Mathews, D. H. ProbKnot: fast prediction of RNA secondary structure including pseudoknots. *RNA* **2010**, *16*, 1870–1880.

(37) Dawson, W. K.; Fujiwara, K.; Kawai, G. Prediction of RNA Pseudoknots Using Heuristic Modeling with Mapping and Sequential Folding. *PLOS ONE* **2007**, *2*, 1–7.

(38) Petingi, L.; Schlick, T. Partitioning and Classification of RNA Secondary Structures into Pseudonotted and Pseudoknot-free Regions Using a Graph-Theoretical Approach. *IAENG Int. J. Comput. Sci.* **2017**, *44*, 241–246.

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

(40) Hajdin, C. E.; Bellaousov, S.; Huggins, W.; Leonard, C. W.; Mathews, D. H.; Weeks, K. M. Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 5498–5503.

(41) Mustoe, A. M.; Lama, N. N.; Irving, P. S.; Olson, S. W.; Weeks, K. M. RNA base-pairing complexity in living cells visualized by correlated chemical probing. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 24574–24582.

(42) Plant, E. P.; Pérez-Alvarado, G. C.; Jacobs, J. L.; Mukhopadhyay, B.; Hennig, M.; Dinman, J. D. A three-stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLOS Biol.* **2005**, *3*, e172.

(43) Trinity, L.; Lansing, L.; Jabbari, H.; Stege, U. SARS-CoV-2 ribosomal frameshifting pseudoknot: Improved secondary structure prediction and detection of inter-viral structural similarity. **2020**, bioRxiv, doi: 10.1101/2020.09.15.298604, preprint posted September 2020.

(44) Omar, S. I.; Zhao, M.; Sekar, R. V.; Moghadam, S. A.; Tuszynski, J. A.; Woodside, M. T. Modeling the structure of the frameshift-stimulatory pseudoknot in SARS-CoV-2 reveals multiple possible conformers. *PLOS Comput. Biol.* **2021**, *17*, e1008603.

(45) Wacker, A.; Weigand, J. E.; Akabayov, S. R.; Altincekic, N.; Bains, J. K., et al. Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **2020**, *48*, 12415–12435.

(46) Rangan, R.; Watkins, A. M.; Chacon, J.; Kladwang, W.; Zheludev, I. N., et al. De novo 3D models of SARS-CoV-2 RNA elements and small-molecule-binding RNAs to aid drug discovery. *Nucleic Acids Res.* **2021**, doi: 10.1093/nar/gkab119.

(47) Ziv, O.; Price, J.; Shalamova, L.; Kamenova, T.; Goodfellow, I.; Weber, F.; Miska, E. A. The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2. *Mol. Cell* **2020**, *80*, 1067–1077.e5.

(48) Smola, M. J.; Rice, G. M.; Busan, S.; Siegfried, N. A.; Weeks, K. M. Selective 2′- hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis. *Nat. Protoc.* **2015**, *10*, 1643–1669.

(49) Ahmed, F.; Sharma, M.; Al-Ghamdi, A. A.; Al-Yami, S. M.; Al-Salami, A. M., et al. A Comprehensive Analysis of cis-Acting RNA Elements in the SARS-CoV-2 Genome by a Bioinformatics Approach. *Front. Genet.* **2020**, *11*, 1385.

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

(51) Iserman, C.; Roden, C. A.; Boerneke, M. A.; Sealfon, R. S.; McLaughlin, G. A.; Jungreis, I.; Fritch, E. J., et al. Genomic RNA Elements Drive Phase Separation of the SARS-CoV-2 Nucleocapsid. *Mol. Cell* **2020**, *80*, 1078–1091.

(52) Reuter, J.; Mathews, D. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinform.* **2010**, *11*, 129.

(53) Yan, S.; Jain, S.; Zhu, Q.; Schlick, T. Length dependent 3D structures and motions of SARS-CoV-2 frameshifting pseudoknot and alternative pseudoknot elements. **2021**, in preparation.

(54) Biesiada, M.; Purzycka, K. J.; Szachniuk, M.; Blazewicz, J.; Adamiak, R. W. Automated RNA 3D Structure Prediction with RNAComposer. *Methods Mol. Biol.* **2016**, *1490*, 199–215.

(55) Xu, X.; Chen, S.-J. Hierarchical Assembly of RNA Three-Dimensional Structures Based on Loop Templates. *J. Phys. Chem.* **2018**, *122*, 5327–5335.

(56) Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P., et al. SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.* **2016**, *44*, e63–e63.

(57) Krokhotin, A.; Houlihan, K.; Dokholyan, N. V. iFoldRNA v2: folding RNA with constraints. *Bioinform.* **2015**, *31*, 2891–2893.

(58) Abraham, M. J.; Murtola, T.; Schulz, R., et al. GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1-2*, 19–25.

(59) Schlick, T. In silico drug discovery targeting the SARS-CoV-2 frameshifting RNA element. **2021**, Joint Mathematics Meetings: AMS Special Session on Advances in Computational Biomedicine II, talk given on January 8, 2021.

(60) Katoh, K.; Misawa, K.; Kuma, K.; Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **2002**, *30*, 3059–3066.

(61) Baranov, P.; Gurvich, O.; Hammer, A.; Gesteland, R.; Atkins, J. RECODE 2003. *Nucleic Acids Res.* **2003**, *31*, 87–89.

(62) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215*, 403–410.

(63) Clamp, M.; Cuff, J.; Searle, S. M.; Barton, G. J. The Jalview Java alignment editor. *Bioinform.* **2004**, *20*, 426–427.

(64) Antczak, M.; Zok, T.; Popenda, M.; Lukasiak, P.; Adamiak, R.; Blazewicz, J.; Szachniuk, M. RNApdbee – a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs. *Nucleic Acids Res.* **2014**, *42*, W368–372.

(65) Wilkinson, K. A.; Merino, E. J.; Weeks, K. M. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): Quantitative RNA structure analysis at single nucleotide resolution. *Nat. Protoc.* **2006**, *1*, 1610–1616.

(66) Busan, S.; Weeks, K. M. Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2. *RNA* **2018**, *24*, 143–148.

(67) Zgarbová, M.; Otyepka, M.; Šponer, J.; Mládek, A.; Banáš, P.; Cheatham, T. E.; Jurečka, P. Refinement of the Cornell et al. Nucleic Acids Force Field Based on Reference Quantum Chemical Calculations of Glycosidic Torsion Profiles. *J. Chem.* **2011**, *7*, 2886–2902.

(68) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(69) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(70) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

Footnotes

* We use the 3_6, 3_3 pseudoknot and 3_5 junction notations throughout as long as the central FSE region contains these independently folded structures.

** Created with BioRender.com