PNAS RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY OPEN ACCESS

Evolution of coronavirus frameshifting elements: Competing stem networks explain conservation and variability

Shuting Yan[a], Qiyao Zhu[b], Jenna Hohl[b], Alex Dong[a], and Tamar Schlick[a,b,c,d,1]

Edited by Joseph Puglisi, Stanford University School of Medicine, Stanford, CA; received December 15, 2022; accepted March 27, 2023

The frameshifting RNA element (FSE) in coronaviruses (CoVs) regulates the programmed −1 ribosomal frameshift (−1 PRF) mechanism common to many viruses. The FSE is of particular interest as a promising drug candidate. Its associated pseudoknot or stem loop structure is thought to play a large role in frameshifting and thus viral protein production. To investigate the FSE structural evolution, we use our graph theory-based methods for representing RNA secondary structures in the RNA-As-Graphs (RAG) framework to calculate conformational landscapes of viral FSEs with increasing sequence lengths for representative 10 Alpha and 13 Beta-CoVs. By following length-dependent conformational changes, we show that FSE sequences encode many possible competing stems which in turn favor certain FSE topologies, including a variety of pseudoknots, stem loops, and junctions. We explain alternative competing stems and topological FSE changes by recurring patterns of mutations. At the same time, FSE topology robustness can be understood by shifted stems within different sequence contexts and base pair coevolution. We further propose that the topology changes reflected by length-dependent conformations contribute to tuning the frameshifting efficiency. Our work provides tools to analyze virus sequence/structure correlations, explains how sequence and FSE structure have evolved for CoVs, and provides insights into potential mutations for therapeutic applications against a broad spectrum of CoV FSEs by targeting key sequence/structural transitions.

conformational landscape | coronavirus frameshifting element | RNA pseudoknot | RNA-As-Graphs | graph theory

Significance

The RNA topology of the frameshifting element in the SARS-CoV-2 viral genome regulates the programmed −1 ribosomal frameshift crucial for virus propagation. Using graph theory-based methods for representing RNA secondary structures in the RNA-As-Graphs (RAG) framework, we calculate conformational landscapes of viral FSEs with increasing sequence lengths for 23 representative Alpha and Beta-coronaviruses to investigate sequence influence and length dependence on virus structural evolution. For different FSE landscapes, a variety of pseudoknots, stem loops, and junctions emerge. We explain alternative competing stems and topology changes by recurring mutations and topology robustness by coevolution. We propose that topology changes contribute to tuning the frameshifting efficiency, and thus pinpointing key FSE locations involved in competing stems defines new potential therapeutic avenues.

The viral agent of COVID-19, namely SARS-CoV-2, and related coronaviruses (CoVs) pose an ongoing public health threat. While SARS-CoV and MERS-CoV appear to have developed from bats, the origin of SARS-CoV-2 is not yet resolved. With the observed damage and continued emergence of new variants of SARS-CoV-2 and related viruses, it is urgent to understand the characteristics of these CoVs, investigate their evolution, and hamper their transmissibility. Developing antiviral therapeutics for a broad spectrum of CoVs will assist the defense against novel CoV diseases and future public health challenges.

CoVs are positive-sense RNA viruses belonging to the order Nidovirales. One mechanism shared by all CoVs is the −1 programmed ribosomal frameshifting (PRF). PRF allows the production of multiple proteins from the same mRNA by shifting the reading frame and controlling the stoichiometric ratio between viral proteins. In CoVs the slippage occurs by 1 nucleotide in the backward direction. The frameshifting process achieves the optimal proportions of viral protein products, where proteins needed in large quantities are usually encoded early in the sequence in an open reading frame ORF1A, and those required in lower quantities are encoded in a later overlapping downstream reading frame (ORF1B) (1). The upstream ORF of CoVs encodes nonstructural proteins. The frameshift-dependent downstream ORF encodes enzymes involved in transcription and replication (2). The frameshift site between ORF1A and ORF1B usually includes two *cis* elements: a slippery 7-nucleotide sequence followed by a downstream RNA structure (a pseudoknot or stem loop), and a 5 to 7-nucleotide spacer between them. The slippery site has the sequence N NNW WWH (N: for any base; W: U or A; H: A, C, or U), and is U UUA AAC for CoVs, promoting the aminoacylated- and peptidyl-tRNAs to detach from the mRNA and re-pair to codons in the −1 frame (3), Fig. 1. The downstream frameshifting stimulator structure, especially the RNA pseudoknot, not only stalls the ribosome but also interacts with various components of the ribosome complex (6, 7).

The −1 PRF process has been proposed to be a viable target for the prevention of viral propagation (2, 8–15). Thus, there is great interest in understanding the frameshifting mechanism for developing therapeutics against CoVs and other viruses like Alphaviruses and HIV that employ frameshifting (16, 17). Besides factors such as tRNA

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: schlick@nyu.edu.

This article contains supporting information online at http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2221324120/-/DCSupplemental.

Published May 8, 2023.

mutations (18), *cis* and *trans* modulators (19), and long-range RNA/RNA interactions (20) that can suppress or modify the frameshifting efficiency, frameshift stimulators achieve their function through pairing of mRNA with rRNA close to the mRNA entrance channel or exit path, interactions of newly synthesized nascent protein chains with the ribosome peptide exit tunnel, and the 5′ or 3′ FSE structures (16). Here, we focus on the 3′ stimulator FSE structure, which has received significant attention, as it not only impacts all associated biological processes but also the evolution of both viruses and hosts (21). FSE 3D structures for SARS-CoV-2 have been characterized by X-ray crystallography, Cryo-EM, and NMR, and investigated by chemical probing (SHAPE-MaP and DMS-MaPseq), small-angle X-ray scattering diffraction, and computational methods (4, 5, 7, 22–31).

While the effects of mutations in the slippery site and pseudoknot motif sequence on frameshifting efficiency have been investigated (3, 27, 32–34), a thorough understanding of how these mutations affect FSE folding and how RNA folding influences the frameshifting process is lacking. Chemical probing, genome-wide scanning, and computational modeling for the pseudoknot region (4, 5, 25, 26) have suggested a dominant H-type pseudoknot for SARS-CoV-2, as found by Cyro-EM and crystallography (7, 22–24), and several competing FSE conformations as a function of sequence lengths. However, a systematic elucidation of the link between FSE sequence, structure, and function is still unknown, due to the structural complexity, plasticity (35) and length-dependent features (36).

To address these limitations, we develop and apply graph-based coarse-graining representations of RNA topologies to investigate FSE conformational landscapes of a broad spectrum of CoVs, including representative Alpha and Beta-CoVs, as a function of sequence length. Our RNA-As-Graph (RAG) machinery represents RNA 2D structures as coarse-grained dual graphs, with double-stranded helices represented by vertices and single-stranded loops represented by edges (29). Dual graphs can handle pseudoknot motifs (intertwined base pairs) (37, 38). The advantage of using graphs is the ability to capture the topology of RNA while allowing for differences in the lengths of stems and loops. Thus, coarse-grained graphs allow us to group many closely related 2D structures in a cluster that shares a common topology and thus identify global features. Graph theory is also useful for partitioning larger dual graphs into subgraphs (39), including pseudoknots, and thus relating different submotifs.

In our prior work, we mapped the conformational landscape of SARS-CoV-2 by modeling coupled to SHAPE experiments (4). We showed that three viable structures of SARS-CoV-2 FSE RNA exist, including 3_6 3-stem pseudoknot supported by Cryo-EM and crystallography (7, 22–24), 3_3 alternative pseudoknot, and 3_5 3-way junction (Fig. 1, *Bottom*). We also designed structure-altering and strengthening mutations by our graph-based inverse-folding software RAG-IF (40), confirmed by experiments, as therapeutic avenues (4), and explored the viability and dynamics of tertiary structures of these FSEs and structural transformations (36). Here, we continue our exploration of FSE conformational landscapes from an evolutionary point of view for a wide range of CoVs of related species as shown in Fig. 2. We focus on pinpointing how FSE structure evolved from sequence mutations. This is accomplished by comparing sequence and structure conservation in various CoVs, especially from eight recurring stems. Our ensemble-based structural modeling coupled to topology-based structural representation by graphs helps present a global picture of the conformational dependence on sequence length and context for CoV FSEs. These results contribute to our understanding of CoV FSE sequence–structure–function relationships and suggest possible therapeutic targets (residues) for regulating frameshifting via gene editing.

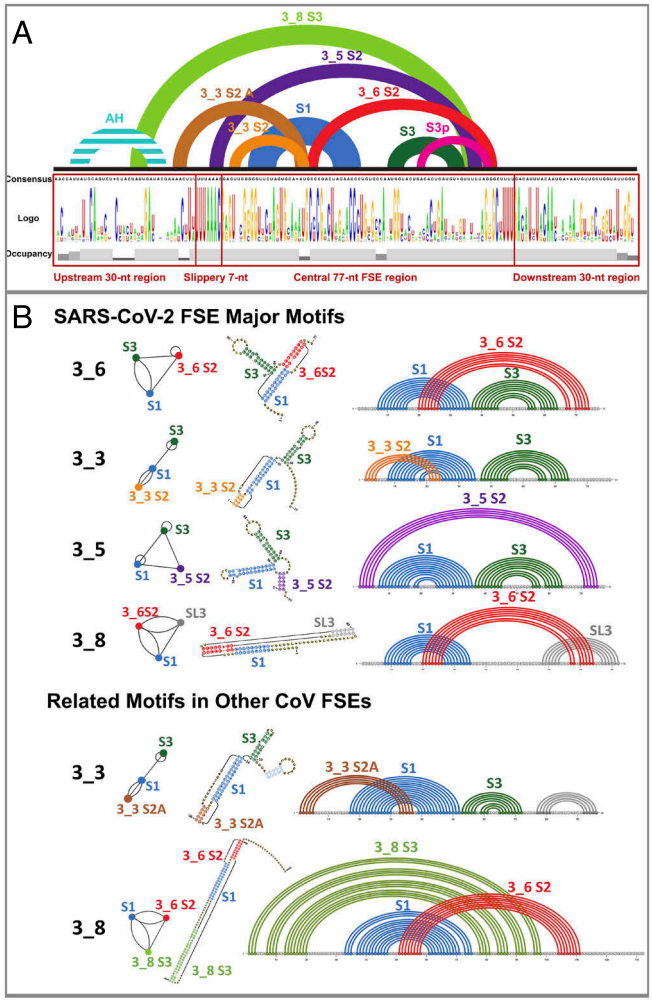

Fig. 1. The common FSE sequence and associated major stems and motifs discussed in this work for Alpha and Beta-CoV FSEs. (*A*) The consensus sequence found by multiple sequence alignment of the 144-nt FSE sequences is written below, with sequence logo and occupancy computed (*SI Appendix* for enlarged view). Due to gap insertions, the consensus sequence is longer than 144 nt. The different FSE regions and the 8 stems in the central FSE region are labeled and aligned accordingly. The attenuator hairpin (AH) upstream of the central FSE region in the Beta-CoV Sarbe subgenus is also shown. (*B*) The SARS-CoV-2 FSE motifs 3_6, 3_3, and 3_5 are derived with SHAPE reactivity data in our prior work (4), and 3_8 was identified as the dominant structure by Huston et al. (5) in their 500-nt windows. FSE motifs involving different stems observed in other CoVs in this work are shown with examples, including Feline FSE 3_3 motif and HCoV-OC43 FSE 3_8 motif.

Results

CoVs are divided into four distinctive genera based on genetic and serological characterization (41). These include Alpha-CoV, Beta-CoV, Gamma-CoV, and Delta-CoV. These groups are thought to have diverged from each other around 2400–3000 BC and infect different groups of animals. Alpha-CoV and Beta-CoV are found mostly in mammals, while Gamma and Delta-CoVs primarily infect birds, although Gamma-CoVs also infect some cetaceans.(41) We select 10 representative Alpha and 13 Beta-CoVs (from each subgenus of families that infect humans) to study the conformational landscape evolution of their FSEs (Fig. 2). Complete RNA genome sequences are downloaded from GenBank. To extract the 144-nt FSE sequences of the 23 CoVs, we locate the 7-nt slippery sites (UUUAAAC), identify the 77-nt

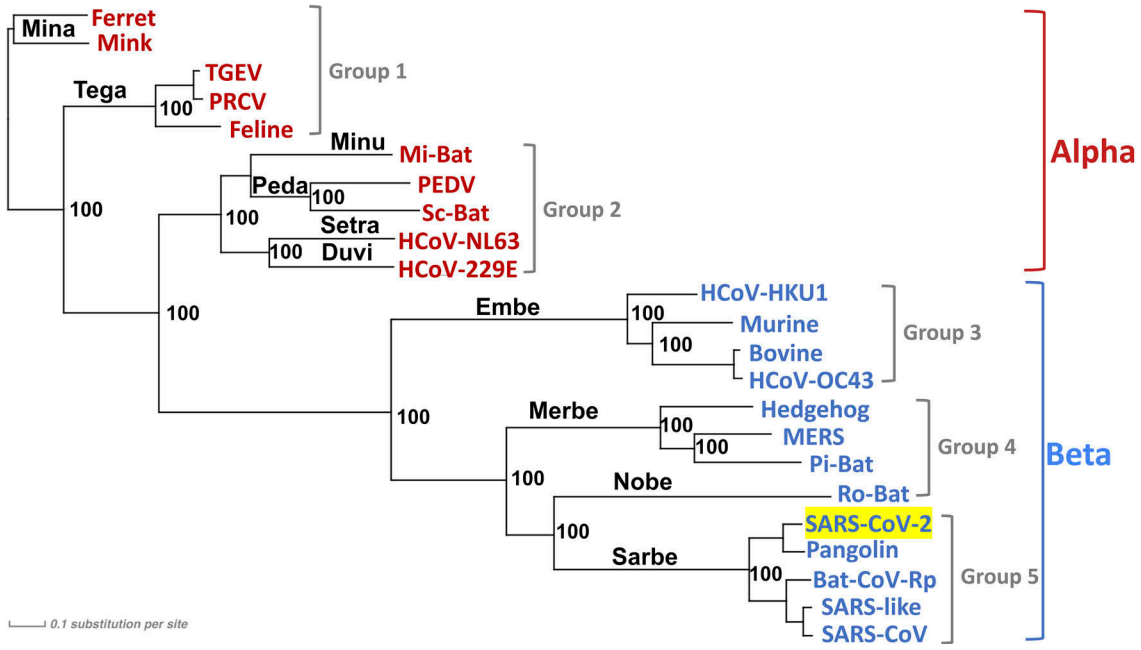

Fig. 2. Phylogenetic tree of 10 Alpha and 13 Beta-CoVs using whole genome sequences. The tree is computed using the general time reversible model (*Methods*). Bootstrap scores (/100) are marked at branch nodes. The scale bar represents 0.1 substitution per nucleotide position, and the branch lengths represent the average substitution numbers per site. The CoVs are numbered from 1 to 23 according to the phylogenetic tree and grouped by features discussed later.

central FSE regions, and add 30 upstream and 30 downstream nucleotides. The 84-nt FSE sequences are listed in *SI Appendix*, Tables S1 for Alpha-CoVs and *SI Appendix*, Table S2 for Beta-CoVs).

Analysis of Topologies from Competing Stems. For each CoV FSE, we perform secondary structure ensemble predictions with free-energy estimates for sequences from 77 to 144 nt, as shown in Fig. 3. Namely, starting from the 77-nt central FSE region,

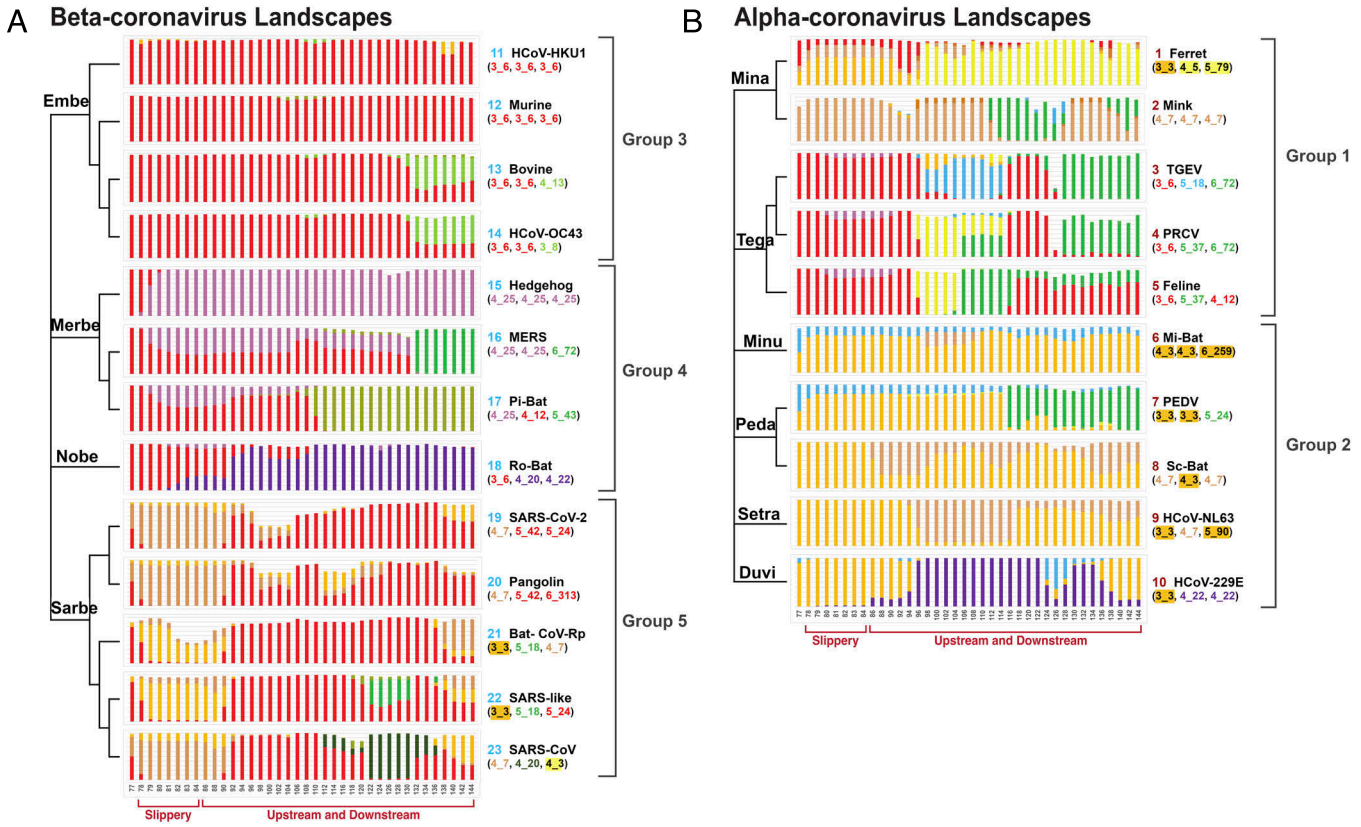

Fig. 3. Conformational landscapes of 23 CoV FSEs. Landscapes are ordered according to the Fig. 2 tree for Beta-CoV and Alpha-CoV in (*A*) and (*B*), respectively. Against each FSE sequence length from 77 to 144 nt, we display bars colored proportionately to probabilities (0 to 100%) of the various FSE motifs. Color codes and dual graphs for motifs are in Fig. 4. Representative dual graph IDs of short, middle, and long sequence lengths are noted following the landscape at right.

we expand to 84 nt by adding the slippery site nucleotides one at a time, after which we add 1 nt on each side until reaching 144 nt. The predicted 2D structures as a function of sequence length from NUPACK (42), are represented as dual graphs. The graph vertices denote stems, and edges are represented by loops. The graph ID is determined by the motif, that is, stems within the central 77-nt FSE region (*Methods* and *SI Appendix*, Fig. S2).

Within the different FSE motifs, we identify eight stems that recur frequently as shown in Fig. 1. Details of the stems in the secondary structures are shown in Figs. 1 and 4, *SI Appendix*, S3–S6 for each Beta-CoV FSE. Stem 1 (blue) is preserved through FSE motifs, but covariations that preserve the base pairing can be found (*SI Appendix*, Fig. S7, which shows MSA of 23 sequences, and consensus structures of Alpha and Beta-CoV FSEs by R-scape (43) with notable covariations indicated). The lengths of stems 1 (blue) and 3 (green) vary according to different FSE contexts. Thus, 3_3 stem 2 and stem 3p emerge and accompany relatively short stems 1 and 3. Besides unknotted motifs with mainly two hairpins from stems 1 and 3, most FSE motifs are three-stem or four-stem pseudoknots/junctions.

As noted above, in our prior work, we identified three FSE structures: 3_6 pseudoknot dominant at short FSE sequences, 3_3 pseudoknot preferred at longer sequences, and the 3_5 3-way junction as a minor conformer (Fig. 1) (4). Together with stem 1 and stem 3, 3_6 stem 2 (red) gives rise to the prevalent H-type pseudoknot for SARS-CoV-2 FSE (Fig. 1) (7, 22). The 3_3 stem 2A (only seen in certain Alpha-CoV FSEs) or 3_3 stem 2 (seen in SARS-CoV-2 3_3) contributes to the alternative pseudoknot 3_3 (Fig. 1). The 3_5 stem 2 leads to the three-way junction. Another pseudoknot, 3_8 motif, contains stem 1, 3_6 stem 2, and 3_8 stem 3 formed by an upstream sequence (Fig. 1). Thus, 3_6 stem 2, 3_3 stem 2, and 3_3 stem 2A compete with each other. Stem 3, 3_5 stem 2, and 3_8 stem 3 also compete, especially for FSEs within the Alpha-CoV Embe subgenus. The competing stems share a common strand, which appears in related sequences (*SI Appendix*, Table S3).

Multiple Sequence Alignment (MSA) Analysis. To further analyze the stem and sequence conservation, we identify the consensus sequence by performing multiple sequence alignment (MSA) and calculate the sequence logo and occupancy (Fig. 1; see Methods and enlarged view in *SI Appendix*). The logo stack heights indicate the sequence conservation, and the occupancy height (area) points to where gaps (low occupancy) occur due to sequence mismatches.

Thus, we see that in addition to the fully conserved slippery site, stem 1 strand regions are highly conserved, consistent with our observation above. While the 3′ end of the central FSE region and the downstream nucleotides also show high conservation, the loop region of stem 1 and the upstream nucleotides are poorly conserved, with notable gaps. Interestingly, these gaps correspond to a) the competing site of 3_6 stem 2, 3_3 stem 2, and 3_3 stem 2A and b) the competing site of the 5′ strand regions of 3_3 stem 2A and 3_8 stem 3.

This finding suggests that frequent mutations within these gaps might trigger alternative stems, namely 3_3 stem 2 and 3_3 stem 2A, and thus play key roles in CoV FSE conformation evolution.

Phylogenetic Analysis of Selected Alpha and Beta-CoVs. We built a maximum-likelihood phylogenetic tree based on the complete RNA genomes of selected CoVs using the general time reversible model (44). The best tree is chosen from the maximum likelihood measure based on the substitution model assuming different rates of substitution and different frequencies of nucleotide occurrence (*Methods*). Overall, the phylogenetic tree shows high confidence in the branch nodes with perfect bootstrap scores (Fig. 2). The Beta-CoVs form well-separated clusters corresponding to the 4 subgenera (Embe-, Merbe-, Nobe-, and Sarbe-covirus), while the Alpha-CoVs are more disperse, especially the Minu-, Peda-, Setra-, and Duvicovirus subgenera. Rather than suggesting the true chronological evolution of CoVs, this tree suggests a path from Alpha to Beta-CoVs via nucleotide substitutions. We then use this path to trace how the FSE conformational landscapes evolved with these mutations.

Overview of CoV FSE Conformational Landscapes. Using the 2D structure ensembles and associated Boltzmann distribution, we compute probabilities of different FSE motifs for each FSE length and plot conformational landscapes for our selected group of 23 representative coronavirus FSEs in Fig. 3*A* for Beta-CoV FSEs and Fig. 3*B* for Alpha-CoV FSEs. A total of 12 major motifs emerge, consisting of different combinations of the eight stems (shown in Fig. 4*A*). Except for two, all motifs contain the more stable stems 1 and 3 (4). Based on addition of other stems, the motifs can be partitioned into 5 basic groups as shown in Fig. 4*A*: 1) unknotted tandem stems 1 and 3 (blue), 2) 3-stem H-type pseudoknot 3_6 (red), 3) 3_5-like (3_5 stem 2 containing) (purple), 4) 3_8 stem 3 containing (green), and 5) 3_3-like (3_3 stem 2 or 3_3 stem 2A containing) (orange). All dual graph representations are shown in Fig. 4*B* for the 12 major FSE motifs. The core subgraph of each motif is highlighted to reflect graph relationships.

Most FSE conformational landscapes in this study exhibit length-dependent variations. Only Beta-CoV HCoV-HKU1, Murine, and Alpha-CoV Mi-Bat FSEs show a stable dominant motif over increasing lengths (Fig. 3*A* Nos. 11 and 12, and Fig. 3*B* No. 6). Adding upstream and downstream sequences contributes as follows: 1) These residues provide a competing strand for pairing, e.g., 3_6 to 3_3 transition with upstream contribution in 3_3 stem 2 and 3_3 to 4_7 transition with downstream residue base pairing in S3p; 2) these residues allow formation of small hairpins beyond the central FSE region, e.g., FSEs from Beta-CoV Embe and Sarbe subgenus (Fig. 3*A* top 4 Nos. 11–14 for Embe and bottom 5 Nos. 19–23 for Sarbe).

With these in mind, we describe the following length-dependent patterns that occur: 1) For some species, a stable topology dominates at all lengths, for example, Beta-CoV Embe (Fig. 3*A* top 4 Nos. 11–14); 2) adding the slippery site introduces alternative structures, for example, Beta-CoV Sarbe subgenus (Fig. 3*A* bottom 5 Nos. 19–23); 3) topological transitions tend to occur as sequence length increases, for example, Beta-CoV Merbe and Nobe (Fig. 3*A* Nos. 15–18); 4) topology temporarily changes and recovers at long sequence contexts, for example, Alpha-CoV Setra and Duvi (Fig. 3*B*, *Bottom* 2 Nos. 9–10).

We see that the 3_6 pseudoknot and 3_3-like motifs are the most prevailing motifs, especially 3_6 pseudoknot for Beta-CoVs. Indeed, these two motif groups were found by SHAPE experiments for SARS-CoV-2 FSE (4). The 3_8 stem 3 motifs appear in 11 coronavirus landscapes for sequences >100 nt, as pairing with the upstream nucleotides is involved. The 3_5-like (3-way junction) motifs are specific to the Duvi subgenus of Alpha-CoVs (No. 10), and the Merbe (Nos. 15–17) and Nobe (No. 18) subgenera of Beta-CoVs. Yet, they dominate several landscapes, suggesting their potential of being a strong competitor to 3_6-like and 3_3-like motifs. Indeed, we have

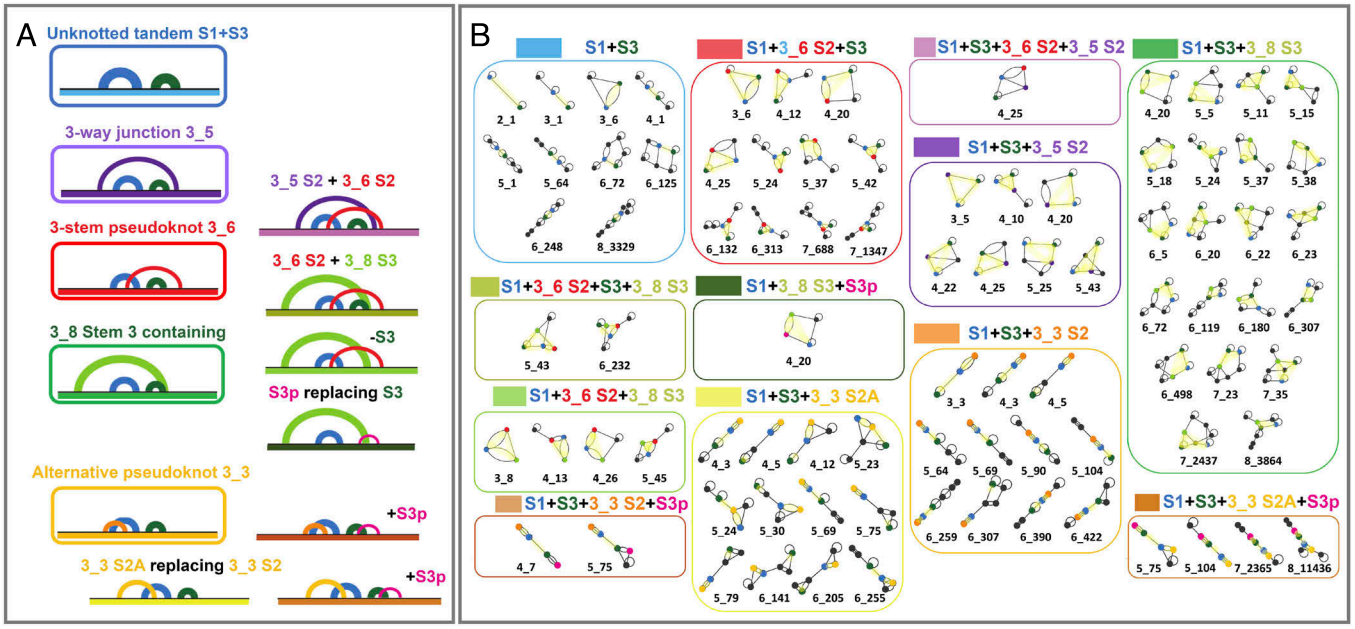

Fig. 4. Motifs defined by the combination of key competing stems. (*A*) The 5 groups of motifs, and 12 major motifs for mapping conformational landscapes. (*B*) Dual graph representations of FSE structures in each of the 12 major motifs. Nodes representing key stems in dual graphs are colored according to the stems in Fig. 1. Shared key subgraphs are highlighted in yellow.

found mere 2-residue mutations that transform a 3_6 wild-type landscape of SARS-CoV-2 FSE into a 3_5 unknotted topology (4, 29).

FSE Motif Emergence as a Result of Evolutionary Mutations.

Overview. Our multiple sequence analysis in Fig. 5 (ordered according to Fig. 2) allows us to study the evolution pattern of FSE conformations shown in Fig. 6 as mutations accumulate and partition the CoVs into 5 groups.

Group 1 viruses, Nos. 1–5 in Fig. 3B. Mina- and Tega-covirus landscapes contain all 5 motif groups and are highly sensitive to sequence lengths. The major conformational changes from 3_3-like motifs of Mina-covirus No. 1-2 to 3_6 pseudoknot of Tega-covirus Nos. 3–5, and 3_8 stem 3 emerge with No. 2 Mink. The sequence alignment in Fig. 5 reveals a shift from 3_3 stem 2A to 3_6 stem 2 due to a G-A mutation in the 3′ strand of 3_6 stem 2 (labeled in red and highlighted in yellow). Fig. 6 box 1 and *SI Appendix*, Fig. S8 show the transition from 3_3 (and 3_3 containing) motif to 3_6 (and 3_6 containing) by mutations (consult Fig. 1) via two major steps: 1) The long stem 3 splits into to a short stem 3 and stem 3p (3_3 to 4_7), and 2) 3_6 stem 2 replaces 3_3 stem 2 and stem 3p. For 3_8 stem 3 to be viable, mutations on both strands form strong AU and GC base pairs.

Group 2 viruses, Nos. 6-10 in Fig. 3B. Remaining Alpha-CoV (Minu, Peda, Setra, and Duvi) landscapes all have 3_3 (formed with 3_3 S2) as major FSE motifs. From Fig. 5, we see that the 3′ strand residue of 3_6 stem 2 continues to mutate from A in Tega-covirus No. 3-5 to U (highlighted in yellow), thereby breaking the AU base pair of 3_6 stem 2. This elimination of 3_6 stem 2 allows 3_3 S2 to dominate the landscape. Within this group, we observe a sudden rise of 3_8 stem 3 in the No. 7 PEDV FSE, due to supporting U-C mutations on both strands. Similarly, the 3_5 stem 2 emergence within the No. 10 HCoV_229E FSE is promoted by a G-C and a C-A mutation on the 5′ strand. The transitions are illustrated in Fig. 6 box 2a-2b and *SI Appendix*, Fig. S9 with related secondary structures and dual graphs. The shift of stem 3 also accompanies the emergence of 3_8 stem 3 or 3_5 stem 2 along with multiple mutations.

Group 3 viruses, Nos. 11–15 in Fig. 3A. Beta-CoV Embe-coviruses have stable FSE landscapes dominated by the 3_6 pseudoknot, especially HCoV-HKU1 and Murine coronavirus (Fig. 3 Nos. 11–12). 3_6 stem 2 is restored due to a 3-nt insertion in the 5′ strand, which extends 3_6 stem 2 from 5 base pairs to 8 (yellow highlight and the red box in the *Top Middle* of Fig. 5). In addition, a U-G mutation in the 3′ strand of 3_6 stem 2 helps form strong consecutive GC base pairs. Within these Embe-coviruses, stem 3 transits to 3_8 stem 3, by an A-U mutation on the 5′ strand of 3_8 stem 3 and a U-A mutation on the common region of the 3′ strands (*SI Appendix*, Fig. S10).

Group 4 viruses, Nos. 15–18 in Fig. 3A. Merbe- and Nobe-covirus FSEs contain competing 3_6 stem 2, 3_5 stem 2, and 3_8 stem 3 (Fig. 3). For 3_5 stem 2 and 3_8 stem 3 to form, stem 1 shortens by a G-A mutation, and stem 3 is shifted upstream by C-G and A-U mutations (highlighted in yellow in Fig. 5). Within this group, 3 supporting mutations for 3_8 stem 3 occur in the No. 16-17 MERS and Pi-Bat FSEs, and 2 supporting mutations for 3_5 stem 2 occur in the No. 18 Ro-Bat FSE. Fig. 6 box 4 and *SI Appendix*, Fig. S11 show that the mutually exclusive 3_5 stem 2 and 3_8 stem 3 can coexist with 3_6 stem 2. However, a long 3_5 stem 2 inhibits the formation of 3_6 stem 2 and 3_8 stem 3. Mutations change the pairing in 3_8 stem 3, which further affects the formation of other stems.

Group 5 viruses, Nos. 19–23 in Fig. 3A. In Sarbe-coviruses, 3_6 stem 2 and 3_3 stem 2 compete (Fig. 3). To inhibit 3_5 stem 2, the stem 1 shortening G-A mutation reverts to G, thus breaking 3_5 stem 2 and supporting 3_3 stem 2 (highlighted in yellow in Fig. 5). The 3-nt insertion in the 5′ strand of 3_6 stem 2 of Embe-coviruses disappears, restoring 3_3 stem 2. While stem 3p accompanies 3_3 stem 2 in most Sarbe-coviruses, it is inhibited in Nos. 21–22 Bat-CoV-Rp and SARS-like coronavirus by a U-C mutation. *SI Appendix*, Fig. S12 shows the formation of stem 3p and 3_8 stem 3 depends on base pairing of stem 3.

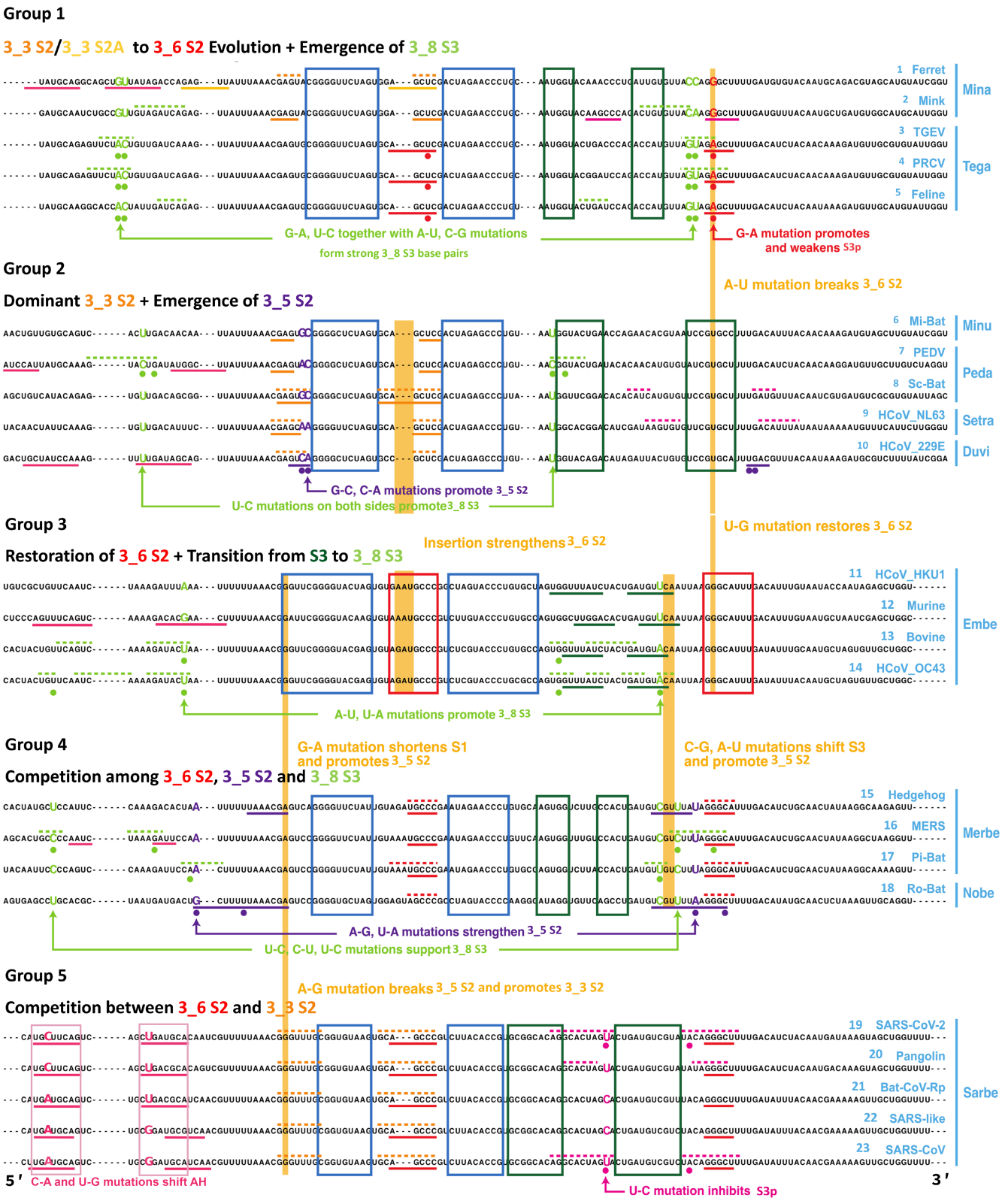

Fig. 5. Multiple sequence alignment and mutation analysis of 23 CoV FSEs. The sequences are ordered according to Fig. 2 and partitioned into 5 groups based on their conformational landscape similarity in Fig. 3. Within each group, the conserved stems are boxed, and the alternating stems are marked, with dominant stems drawn below as solid lines and the minor stems marked above by dashed lines. The conformational transition mutations are written in the associated stems' colors, and dots below show base pairs. For neighboring groups, conformational evolution mutations are highlighted in orange. AHs are identified by the red underline/box at the 5′-end in some CoV FSEs, including HCoV-229E, MERS, and all five sequences in the Sarbe subgenus.

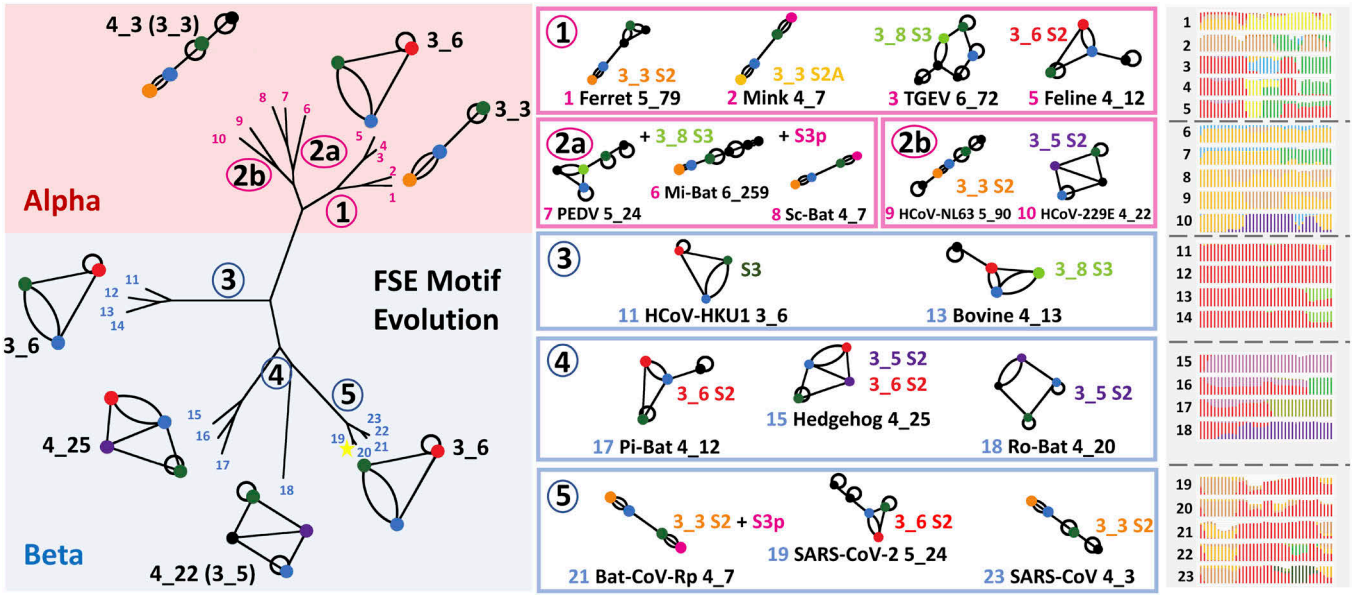

Fig. 6. Motif transition path along the phylogenetic tree and key transitions identified in each group of FSEs.

Summary and Discussion

By analyzing FSEs for 23 representative Alpha and Beta-CoVs as a function of increasing sequence length, mimicking ribosomal translation, we have traced and related FSE motif changes to sequence alterations in evolutionary trees. In Fig. 6, we summarize the FSE motif evolution tree and identify major motif transition paths among different CoV subgenera. Our conformational landscape tool helps dissect coronavirus FSE evolution and interpret alternative conformations observed for SARS-CoV-2 (4). Because motif changes likely relate to frameshifting efficiency, this combined structural biophysics and evolutionary point of view is therapeutically important. The mutations in Table 1 and Fig. 7 based on our competing stem analysis define good targets for altering or stabilizing FSE conformations by gene editing or drug binding.

5′ and 3′ Stimulators of Ribosomal Frameshift Sites. The 3′ stimulator structures in different viruses vary from simple stem loops to more elaborate stem loop, compact pseudoknots, larger pseudoknots, and apical loop–internal loop complexes (ALIL pseudoknot). (16) Frameshifting is affected by multiple features of 3′ intra-mRNA structural stimulators, including structural plasticity, thermodynamic stability, and torsional resistance. Among various structural features, stem and loop lengths and specific three-dimensional structures play important roles.

From prior predictions, the FSE of murine CoV MHV-A59 is a 2-stem pseudoknot (45), similar to that of the avian coronavirus infectious bronchitis virus (IBV) (46) (dual graph: 2_3). The FSE of HCoV-229E (our Alpha-CoV No. 10 in Fig. 3) is an "elaborated" pseudoknot of three stems involving a long-range kissing hairpin (47) (dual graph: 3_8). The putative FSE structures of PEDV (48) (Alpha-CoV No. 7) and TGEV (49) (Alpha-CoV No. 3) are predicted to adopt a similar conformation (dual graph: 3_8) for HCoV-229E FSE. Although the HCoV-229E landscape in our work is dominated by 3_3 or 3_5 containing motifs, S1 is conserved, and Loop 1 residues are unpaired. The S2 5′-strand in a previous study (47) is single-stranded in 3_5 for middle-length sequences (e.g., 100 nt), base-paired within 3_3 motif for long sequences (e.g., 144 nt), and unpaired in extended sequences (e.g., 184 nt).

For the SARS-CoV-2 FSE, different modeling and experimental conditions have suggested different folds for various lengths. The attenuator hairpin (AH) upstream of the slippery site can alter the −1 frameshifting activity in SARS-CoV (50). Including the 5′ stimulator AH introduces more variations in the FSE structure. Due to low sequence conservation in the AH region, only a few Alpha-CoV FSE adopt AH, while all 5 FSE of the Beta-CoV Sarbe subgenus plus MERS have the 5′ stimulator AH (Fig. 5, Group 5 and MERS from Group 4).

Our combined structural findings for SARS-CoV-2 include the following themes (*SI Appendix*, Fig. S13 for a comparison

Table 1. Proposed FSE mutants designs inspired by this work's evolutionary analysis

Design		Mutant	Function	Base pairs involving mutations
RAG-IF graph-based inverse folding	(a)	[U41C, U101C, U103A, C105A]	4_7 to 3_3	C-G formed by U41C, U103A, and C105A break base pairs
	(b)	[A102U, U103C, A104G, C105U]	3_6 to 4_7	A102U, U103C, A104G form base pairs
	(c)	[G100A, U101G, U103A, A104U, C105U]	3_6 to 3_3	U103A, A104U, C105U form base pairs
	(d)	[C27G, G28U, G39U, G58C, C60G, A102U, C105A]	3_6 to 3_5	G58C breaks 3_6 S2 pairing, C27G, G28U, A102U, and C105A form 3_5 S2 pairing
Manual	(e)	[AUC25-27del, G28C, G39A, C105A]	strengthen 3_6	G39A weakens 3_3 S2 in 3_3
	(f)	[AUC25-27del, A24G, G28C, G39A]	strengthen 3_6	G39A weakens 3_3 S2 in 3_3
	(g)	[AUC25-27del, A24G, G28C, C105A]	3_6 to 3_3	Slippery site residues unpaired due to deletions

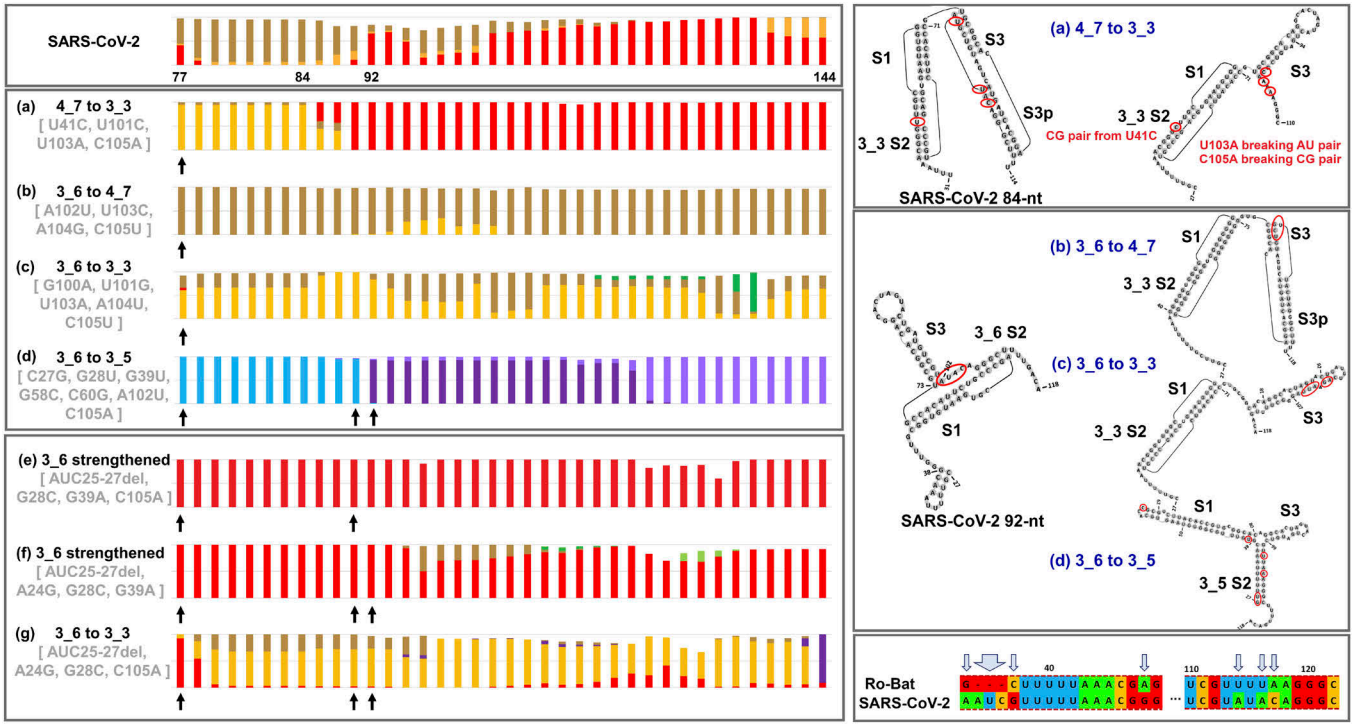

Fig. 7. Modified landscapes resulting from proposed SARS-CoV-2 mutations (Table 1) deduced from conformational landscapes (Fig. 3) and probed by RAG-IF experiments in this work.

of our results here to other works (5, 25, 51)): 1) dominant 3-stem pseudoknot (dual graph: 3_6), in agreement with X-ray crystallography (22, 23), NMR (28), SAXS (27), and Cryo-EM (7, 24) at short constructs, which can also coexist with AH and AS1 at mid-length sequences(4, 5, 51); 2) length-dependent pseudoknots and 3-way junctions at short- and middle-length sequence by SHAPE and ShapeKnots predictions (4) (dual graphs: 3_6, 3_3, and 3_5), with the minor 3-way junction also supported by DMS reactivity (51); 3) unknotted stem loop structure containing AH and a long AS1 that replaces S1 at long sequences by DMS-MaPseq and DREEM clustering (25, 51) (dual graph: 2_2; *SI Appendix*, Fig. S13); 3) different pseudoknot with same 3_6 S2 and a different S3 formed with downstream sequences together with AH and AS1 at long sequences supported by in vivo SHAPE-MaP and ShapeKnots predictions (5) (dual graph: 3_8; *SI Appendix*, Fig. S13). Thus, all the 2- and 3-stem topologies 2_2, 3_6, 3_3, 3_8, and 3_5, of which the middle 3 motifs are pseudoknots, appear to be viable FSE structures in the complex landscape of SARS-CoV-2.

Critical Mutations Switching Stem Formations. The length-dependent features of our FSE conformational landscape reveal a complex network of competing stems, where parts of all the possible stems combine and lead to many types of RNA graph motifs (Fig. 1). The potential of various stems to form results from the recurrence of short sequence fragments, which can pair with the same common strand. Wobble base pairs together with bulges and internal loops in the helix increase the tolerance of sequence variation, including point mutations and insertions/deletions. The gaps in multiple sequence alignment correspond to insertions/deletions, which distinguish different subgenera and even families of CoVs. Key insertions/deletions can enhance a certain stem, leading to a dominant motif. For example, the existence of the 3_3 stem 2A in some Alpha-CoV FSEs can be explained by insertions in Beta-CoV FSEs at positions of its 3′ strand (Embe, Merbe, and Nobe) or 5′ strand (Sarbe) that break the pairing of 3_3 stem 2A. Another example is the enhanced stability of 3_6 stem 2 through more inserted residues, explaining the dominant 3-stem pseudoknot motif for Beta-CoV FSEs.

The underlying potential of forming various stems from sequence immediately suggests mutations at critical positions to dramatically change the FSE topology that we have confirmed computationally (Table 1 and Fig. 7). These positions in Table 1 include loop or stem closing pairs and residues between two short helices. For short stems, such as 3_5 stem 2 and 3_6 stem 2, a point mutation might inhibit their formations. Large changes of dominant motifs introduced by insertions/deletions correspond well to genera and subgenera classifications, underscoring the utility of coarse-grained graph motif representations of viral functional elements to suggest structural evolution patterns.

Note that many 2D structures correspond to a group topology, and thus, our more global view facilitates pattern identification. In Table 1, mutations (e)–(g) for SARS-CoV-2 are based on nearby Beta-CoV sequences/landscape, and mutations (a)–(d) were achieved by our graph-based inverse folding program RAG-IF (40). Our RAG-IF machinery confirmed the minimal mutations required to either strengthen/change the dominant motif at about 95% at a given sequence length or support certain motifs/stems to dominate the whole landscape. The dominant motif 4_7 in short SARS-CoV-2 FSE sequences can be switched to simpler motifs 3_3 and 2_1. The dominant motif 3_6 for middle and long sequences can be modified into different motifs 3_3, 4_7, and 3_5 in mutations (a)–(d). The 3-nt insertions unique to the Sarbe subgenus support base pairs involving slippery site residues, which might relate to frameshifting and thus can be a potential therapeutic target. The key GA mutation observed in Fig. 5 together with the codon deletion as observed in mutations (e)–(g) lead to the formation of a whole landscape dominated by 3_6 or 3_3 motif.

Topological Conservation of CoV FSEs. Although FSE conformational landscapes exhibit sequence length and context-dependent variations, FSEs maintain topological conservation with tolerance to sequence changes. Covariant base pairs reflect the constraints applied by RNA structure on sequence evolution. These covariations might also distinguish different subgenera or genera, while preserving the helix base pairing. The same graph motifs can be adopted by multiple FSEs within a subfamily, for example, 3_5 stem 2 and 3_8 stem 3 containing motifs in Beta-CoV Merbe FSEs. With mutations, FSEs adapt to form the same stems with distinct bases among three viruses in this group (Hedgehog, MERS, and Pi-Bat), while the stem length varies and actual pairings are nonhomologous. Displacements of stems indicate the evolution of RNA secondary structure from their underlying coding sequences. Our coarse-grained graph representation of RNA structures can highlight aspects of topology conservation that originate from displaced pairs and stem length adjustment. FSEs also tend to preserve their topologies by splitting a long stem into two intertwined short stems and leaving other stems unchanged. A case in point is the emergence of 3_3 stem 2 from stem 1, or stem 3p from stem 3.

Evolution and Function. Our FSE conformational landscapes suggest information on dominant motifs as well as highlight the functional importance of structural plasticity. Beta-CoV Sarbe FSEs experience a motif change from 3-stem pseudoknot 3_6 to 3_3 alternative pseudoknot when more slippery site residues are added, and the 3_6 pseudoknot prevails after all slippery site residues are included. Our previous work reveals that a short flanking stem aids in the formation of the 3_3 alternative pseudoknot for SARS-CoV-2 (4). This stem exists temporarily and is replaced by small hairpins upstream/downstream beyond the central FSE region. This phenomenon recurs consistently among Sarbe FSEs but is less obvious in other Beta-CoV FSEs.

Similar changes also appear for Alpha-CoV HCoV-229E where the 3_3 pseudoknot is replaced by the 3-way junction FSE 3_5 after addition of the slippery site and Alpha-CoV HCoV-NL63 where 3_3 topology switches to 4_7 with the addition of stem 3p.

Conformational changes triggered by the addition of slippery site residues might be important due to the tRNA translocation during frameshifting. Different FSE folding might introduce various resistance to unfolding and consequently generate different mechanical tension on the viral RNA. Ribosome traffic/collisions can also affect frameshifting (52).

In addition to FSE RNA structural changes, evolution leads to phylogenetic specificity in the overlap of ORF1A and ORF1B (53) and the frame 0 stop codon location. The frame 0 stop codon of FSEs is located 5 codons (in Beta-CoV Sarbe), 7 codons (in Beta-CoV Embe, Merbe, and Nobe), and 10 codons (in Alpha-CoVs) downstream of the frameshift site. This codon is a constituent of stem 1. Two mutations define the stop codon of Sarbe FSEs, and insertions in other Beta-CoV (Embe, Merbe, and Nobe) FSEs contribute to their stop codons. The proximal stop codon position in relation to the frameshifting site is critical for maintaining optimal frameshift efficiency (7). The stop codon of Alpha-CoV FSEs and Beta-CoV Sarbe FSEs is a constituent of stem 1, while that of other Beta-CoV FSEs (Embe, Merbe, and Nobe) is part of 3_6 stem 2.

Limitations and Outlook. Our FSE conformational landscapes provide insight into FSE structural plasticity, especially the length-dependent switch of FSE dominant topologies, which might tune the frameshifting efficiency. In our current study, FSE conformational landscapes rely on secondary structure predictions by NUPACK (54), which can predict pseudoknots, compute energies for alternative conformations, and is computationally feasible for the lengths analyzed here for many sequences. In our previous work, NUPACK landscapes of SARS-CoV-2 FSE were supported by SHAPE reactivity data for selected lengths. The mutations we propose here and viral landscapes can be similarly tested and explored by chemical probing and other biophysical techniques.

As is well known, no single program for 2D structure prediction is perfect, but their application to a large body of related sequences as done here is meaningful. Certainly, free states of FSE RNAs also fold differently from systems where the FSE is wedged in the mRNA channel. Whether the binding of tRNA affects FSE folding is also unknown. FSE conformations at greater sequence lengths containing the attenuator hairpin are yet to be investigated.

Despite these well-recognized limitations, graph theory–based computational modeling of CoV FSE landscapes as used here in combination with biophysical modeling of tertiary FSE structures (4, 36) has expanded our understanding of CoV FSE structure evolution. By combining phylogenetic evolution with conformational motifs points of view, we could pinpoint topological changes brought about by key mutations. These include insertions/deletions between subgenera and structural evolution by shifted stems favored within subgenera (Figs. 3, 5, 6). Besides many intriguing mechanistic insights, our approach can be further applied to predict mutations that lead to dramatic conformational changes and/or result in dominant structures that potentially enhance or diminish frameshifting, as therapeutic avenues. The suggested mutations in Table 1 and other designs, which are proposed to alter SARS-CoV-2 landscape in different ways as shown in Fig. 7, may help obtain more complex/bulky or simpler FSEs. In turn, changes in FSE structures and associated fluctuations may have profound effects on frameshifting efficiencies and viral protein production. Ongoing work in this direction demonstrates the utility of our landscape and mutation analyses.

Materials and Methods

Coronavirus RNA Genome Sequences. The 10 Alpha and 13 Beta-CoV sequences with complete genome coverage are downloaded from the GenBank https://www.ncbi.nlm.nih.gov/genbank/ (accessed 1 July 2022). The accession codes are LC215871 for Ferret coronavirus, MN535736 for Mink coronavirus, EU074218 for Transmissible gastroenteritis virus (TGEV), KY406735 for Porcine respiratory coronavirus (PRCV), EU186072 for Feline coronavirus, NC_010438 for Miniopterus bat coronavirus (Mi-Bat), MK841494 for Porcine epidemic diarrhea virus (PEDV), NC_009657 for Scotophilus bat coronavirus (Sc-Bat), KY983586 for Human coronavirus NL63, KY369909 for Human coronavirus 229E, NC_006577 for Human coronavirus HKU1, NC_048217 for Murine hepatitis virus, NC_003045 for Bovine coronavirus, NC_006213 for Human coronavirus OC43, NC_039207 for Betacoronavirus Erinaceus (Hedgehog), NC_019843 for Middle East respiratory syndrome-related coronavirus (MERS), NC_009020 for Pipistrellus bat coronavirus (Pi-Bat), NC_030886 for Rousettus bat coronavirus (Ro-Bat), NC_045512 for Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), MT121216 for Pangolin coronavirus, JX993987 for Bat coronavirus Rp, KF367457 for Bat SARS-like coronavirus WIV1, NC_004718 for Severe acute respiratory syndrome coronavirus (SARS-CoV).

Coronavirus Frameshifting Regions. Using the *pairwise Alignment* function from the *Biostrings* v2.62.0 (55) package in RStudio, we locate the FSE slippery sites (UUUAAAC) for our 23 coronavirus sequences. We extract the 144-nt FSEs starting from 30-nt upstream of the slippery sites. The 144-nt FSE regions are residues 12165-12308 for Ferret, residues 12289-12432 for Mink, residues

12302-12445 for TGEV, residues 12290-12433 for PRCV, residues 12341-12484 for Feline, residues 12899-13042 for Mi-Bat, residues 12580-12723 for PEDV, residues 12614-12757 for Sc-Bat, residues 12357-12500 for HCoV-NL63, residues 12398-12541 for HCoV-229E, residues 13564-13707 for HCoV-HKU1, residues 13566-13709 for Murine, residues 13305-13448 for Bovine, residues 13304-13447 for HCoV-OC43, residues 13548-13691 for Hedgehog, residues 13397-13540 for MERS, residues 13649-13792 for Pi-Bat, residues 13047-13190 for Ro-Bat, residues 13432-13575 for SARS-CoV-2, residues 13291-13434 for Pangolin, residues 13264-13407 for Bat-CoV-Rp, residues 13362-13505 for SARS-like, and residues 13362-13505 for SARS-CoV.

Multiple Sequence Alignment. We perform multiple sequence alignment (MSA) for both the whole genomes and the 144-nt FSE regions of the 23 CoVs (Fig. 1 and *SI Appendix*, Fig. S1), using the default option of *mafft* v7.505 (56). The FSE MSA is visualized and analyzed using *JalView* v2.11.2.2 (57).

The sequence logo calculated by *JalView* consists of stacks of symbols (A, U, C, and G for RNA), where the stack heights indicate the sequence conservation at each position. Specifically, stack height H_i at position i is computed as

$$H_i = \log_2 N - \left(- \sum_{n=1}^{N} p_{n,i} \log_2 p_{n,i} \right),$$

with N being the total number of distinct symbols (4 for RNA) and $p_{n,i}$ being the frequency of symbol n at position i. The height of each symbol $h_{n,i}$ within the stack shows the relative frequency,

$$h_{n,i} = H_i \cdot p_{n,i}.$$

Phylogenetic Tree. A phylogenetic tree is built upon the whole genome MSA using *RAxML* v8.2.12 (58). The general time reversible (GTR) model is used with optimization of substitution rates and Γ model of rate heterogeneity, by choosing the option *-m GTRGAMMA*. Best-scoring maximum-likelihood tree search and 100 rapid bootstrap replicates are performed in one program run, by choosing the option *-f a* and *-# 100*. Random number seeds for both bootstrap analysis (option $-x$) and parsimony inferences (option $-p$) are chosen to be 12345. The produced phylogenetic tree is visualized using *Dendroscope* v3.8.3 (59).

Dual Graph Representation. Our dual graphs represent RNA secondary structures as undirected and unweighted graphs, where base-paired stems are vertices and single-stranded loops (internal loops/bulges/junctions/pseudoknot loops) are edges (38). In particular, hairpins are denoted as self-loops; dangling 5′ and 3′ ends are ignored; 1-nt internal loops/bulges are ignored; an isolated single base pair is ignored (38). In this work, we update the definition by providing a user-defined tolerance option T for internal loops/bulges, i.e., internal loops/bulges having $\leq T$ single nucleotides on both sides are ignored. In this way, the stems can tolerate larger internal loops/bulges if needed, without being separated into two stems. Here, we set $T = 5$.

Conformational Landscapes. For each coronavirus FSE, we use NUPACK v3.2.2 (42) *subopt* mode and option *-pseudo* to predict RNA secondary structure ensembles for lengths from 77 to 144 nt. The output contains the predicted 2D structures together with their free-energy estimates. We then use the Boltzmann distribution to compute the partition function Z and the probability p_i for each 2D structure i,

$$Z = \sum_{i=1}^{N} e^{-E_i/(k_B \mathrm{T})}, \quad p_i = \frac{e^{-E_i/(k_B \mathrm{T})}}{Z},$$

where E_i is the free-energy estimate of structure i, k_B is the Boltzmann constant, and T is the room temperature (37 °C).

For each length, we apply our dual graph representation and sum up probabilities of structures that correspond to the same graph. Only graphs with probabilities $\geq 1\%$ are retained, and representative 2D structures are recorded for these graphs.

Next, we identify the minimal motifs within these example structures that involve the central 77-nt FSE regions. Our dual graph algorithm tells the residues included in the 5′ and 3′ strands of the stems. We number the stems from the 5′ to 3′ end, and the two strands of a stem are assigned with the same number. For example, in *SI Appendix*, Fig. S2, the strand order of the 2D structure for dual graph 4_20 is 12233414. With these inputs, we use the following algorithm to identify the minimal motif using the algorithm in *SI Appendix*.

For illustration, in *SI Appendix*, Fig. S2, strand order SO for the 2D structures of dual 4_20 is 12233414. Stems 1, 2, 3, and 4 bind with the central FSE region, so $S_{FSE} = [1, 2, 3, 4]$. Since all stems are included, the minimal motif is 12233414.

A slightly more complicated example is the 2D structure for dual 6_186 in *SI Appendix*, Fig. S2. Here, $SO = 112334452566$ and $S_{FSE} = [2, 3, 4, 5]$. Indices of strands with labels in S_{FSE} are 3, 4, 5, 6, 7, 8, 9, and 10, so $LS = 3$, $RS = 10$, and $S_{Motif} = [2, 3, 3, 4, 4, 5, 2, 5]$. Indices of strands with labels in S_{Motif} are 3, 4, 5, 6, 7, 8, 9, and 10, so $LS = 3$, $RS = 10$, and $S_{new} = [2, 3, 3, 4, 4, 5, 2, 5] = S_{Motif}$. Therefore, the update stops, and the minimal motif is 23344525. Renumbering the strands from 1 gives 12233414, which is the same motif as that of dual 4_20.

Once we identify all minimal motifs, we sum up probabilities of dual graphs that correspond to the same minimal motif and retain only motifs with probabilities $\geq 5\%$. These motifs are further grouped manually based on motifs of the 8 recurring stems.

RAG-IF for Minimal Mutations. We use our inverse-folding protocol RAG-IF (40) to mutate the 92-nt SARS-CoV-2 FSE sequence to strengthen/change dominant dual-graph motifs. The sequence length of 92 nt is selected due to a sudden dominant motif change for Sarbe FSEs including SARS-CoV-2. Our RAG-IF program provides minimal mutations by a computational pipeline for RNA design (40, 60), as detailed previously (29). Candidate residues for mutations are selected manually based on the variations from the FSE evolution.

Covariation Analysis. We use Infernal (61) cmbuild and cmcalibrate to build the covariance model from the MSA of the 23 CoV FSE sequences (144 nt) shown in Figs. 1 and 5. FSE homologs are identified via cmsearch from more Alpha- and Beta-CoV sequences to improve the statistics for the covariation analysis, including 1274 Alpha-CoV, 1320 Beta-CoV (Sarbe subgenus excluded), and 73 Beta-CoV Sarbe sequences. The covariation analysis in *SI Appendix*, Fig. S7 is performed via R-scape(43) using the MSA from cmsearch as input, and the consensus structure is predicted via ——cacofold option. The best consensus structure includes the largest possible number of significantly covarying pairs. The maximum-covariation consensus structures are annotated with the significantly covarying base pairs colored in *SI Appendix*, Fig. S7. Base pairs with covariation scores equal or below the target E-value (0.05) are depicted in green. R-scape annotation to depict the alignment positions is shown in the legend.

Data, Materials, and Software Availability. All study data are included in the article and/or *SI Appendix*. Our graph theory-based RNA motif and conformational landscape code and scripts are available on https://github.com/Schlicklab/Coronavirus-FSE-Conformational-Landscape. Coronavirus FSE motif summary, multiple sequence alignment, covariation analysis and mutation design data are also deposited in the same GitHub repository.

ACKNOWLEDGMENTS. We gratefully acknowledge funding from the NSF RAPID Award 2030377 from the Division of Mathematical Sciences and the Division of Chemistry, NSF Division of Mathematics Sciences Award DMS-2151777, NIH R35GM122562 Award from the National Institute of General Medical Sciences, and Philip-Morris International to T. Schlick.

Author affiliations: [a]Department of Chemistry, New York University, New York, NY 10003; [b]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; [c]New York University - East China Normal University Center for Computational Chemistry, New York University Shanghai, Shanghai 200062, China; and [d]New York University Simons Center for Computational Physical Chemistry, New York University, New York, NY 10003

Author contributions: T.S. designed research; S.Y., Q.Z., J.H., A.D., and T.S. performed research; Q.Z. and T.S. contributed new reagents/analytic tools; S.Y., Q.Z., J.H., A.D., and T.S. analyzed data; and S.Y., Q.Z., and T.S. wrote the paper.

1. M. M. Yordanova, P. V. Baranov, Viruses: A frameshift in time. *eLife* **11**, e78373 (2022).
2. E. P. Plant, R. Rakauskaite, D. R. Taylor, J. D. Dinman, Achieving a golden mean: Mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J. Virol.* **84**, 4330–4340 (2010).
3. E. P. Plant, J. D. Dinman, Comparative study of the effects of heptameric slippery site composition on −1 frameshifting among different eukaryotic systems. *RNA* **12**, 666–673 (2006).
4. T. Schlick *et al*., To knot or not to knot: Multiple conformations of the SARS-CoV-2 frameshifting RNA element. *J. Am. Chem. Soc.* **143**, 11404–11422 (2021).
5. N. C. Huston *et al*., Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Mol. Cell* **81**, 584–598.e5 (2021).
6. O. Namy, S. J. Moran, D. I. Stuart, R. J. C. Gilbert, I. Brierley, A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**, 244–247 (2006).
7. P. R. Bhatt *et al*., Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).
8. R. J. Marcheschi, M. Tonelli, A. Kumar, S. E. Butcher, Structure of the HIV-1 frameshift site RNA bound to a small molecule inhibitor of viral replication. *ACS Chem. Biol.* **6**, 857–864 (2011).
9. H. S. Haniff *et al*., Targeting the SARS-CoV-2 RNA genome with small molecule binders and ribonuclease targeting chimera (RIBOTAC) degraders. *ACS Cent. Sci.* **6**, 1713–1721 (2020).
10. Yu. Sun *et al*., Restriction of SARS-CoV-2 replication by targeting programmed −1 ribosomal frameshifting. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2023051118 (2021).
11. S. Munshi *et al*., Identifying inhibitors of −1 programmed ribosomal frameshifting in a broad spectrum of coronaviruses. *Viruses* **14**, 177 (2022).
12. Px. Ren *et al*., A multi-targeting drug design strategy for identifying potent anti-SARS-CoV-2 inhibitors. *Acta Pharmacol. Sinica* **43**, 483–493 (2022).
13. Y. Chen *et al*., A drug screening toolkit based on the 1 ribosomal frameshifting of SARS-CoV-2. *Heliyon* **6**, e04793 (2020).
14. I. Szczesniak, A. Baliga-Gil, A. Jarmolowicz, M. Soszynska-Jozwiak, E. Kierzek, Structural and functional RNA motifs of SARS-CoV-2 and Influenza A virus as a target of viral inhibitors. *Int. J. Mol. Sci.* **24**, 1232 (2023).
15. S. Sreeramulu *et al*., Exploring the druggability of conserved RNA regulatory elements in the SARS-CoV-2 genome. *Angew. Chem. Int. Ed.* **60**, 19191–19200 (2021).
16. J. F. Atkins, G. Loughran, P. R. Bhatt, A. E. Firth, P. V. Baranov, Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–7078 (2016).
17. M. Mikl, Y. Pilpel, E. Segal, High-throughput interrogation of programmed ribosomal frameshifting in human cells. *Nat. Commun.* **11**, 3061 (2020).
18. J. F. Atkins, G. R. Björk, A gripping tale of ribosomal frameshifting: Extragenic suppressors of frameshift mutations spotlight P-site realignment. *Microbiol. Mol. Biol. Rev.* **73**, 178–210 (2009).
19. W. D. Penn, H. R. Harrington, J. P. Schlebach, S. Mukhopadhyay, Regulators of viral frameshifting: More than RNA influences translation events. *Annu. Rev. Virol.* **7**, 219–238 (2020).
20. O. Ziv *et al*., The short- and long-range RNA-RNA interactome of SARS-CoV-2. *Mol. Cell* **80**, 1067–1077.e5 (2020).
21. R. P. Smyth, M. Negroni, A. M. Lever, J. Mak, J. C. Kenyon, RNA structure–A neglected puppet master for the evolution of virus and host immunity. *Front. Immunol.* **9**, 2097 (2018).
22. C. Roman, A. Lewicka, D. Koirala, N. Li, J. Piccirilli, The SARS-CoV-2 programmed −1 ribosomal frameshifting element crystal structure solved to 2.09 Å using chaperone-assisted RNA crystallography. *ACS Chem. Biol.* **16**, 1469–1481 (2021).
23. C. P. Jones, A. R. Ferré-D'Amaré, Crystal structure of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) frameshifting pseudoknot. *RNA (Cambridge)* **28**, 239–249 (2022).
24. K. Zhang *et al*., Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome. *Nat. Struct. Mol. Biol.* **28**, 747–754 (2021).
25. T. C. T. Lan *et al*., Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.* **13**, 1128 (2022).
26. I. Manfredonia *et al*., Genome-wide mapping of SARS-CoV-2 RNA structures identifies therapeutically-relevant elements. *Nucleic Acids Res.* **48**, 12436–12452 (2020).
27. J. A. Kelly *et al*., Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2). *J. Biol. Chem.* **295**, 10741–10748 (2020).
28. A. Wacker *et al*., Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **48**, 12415–12435 (2020).
29. T. Schlick, Q. Zhu, S. Jain, S. Yan, Structure-altering mutations of the SARS-CoV-2 frameshifting RNA element. *Biophys. J.* **120**, 1040–1053 (2021).
30. S. I. Omar *et al*., Modeling the structure of the frameshift-stimulatory pseudoknot in SARS-CoV-2 reveals multiple possible conformers. *PLoS Comput. Biol.* **17**, e1008603 (2021).
31. R. Rangan *et al*., De novo 3D models of SARS-CoV-2 RNA elements from consensus experimental secondary structures. *Nucleic Acids Res.* **49**, 3092–3108 (2021).
32. I. Brierley, A. J. Jenner, S. C. Inglis, Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227**, 463–479 (1992).
33. I. Brierley, N. J. Rolley, A. J. Jenner, S. C. Inglis, Mutational analysis of the RNA pseudoknot component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **220**, 889–902 (1991).
34. E. P. Plant, A. C. Sims, R. S. Baric, J. D. Dinman, D. R. Taylor, Altering SARS Coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses* **5**, 279–294 (2013).
35. C. Moomau, S. Musalgaonkar, Y. A. Khan, J. E. Jones, J. D. Dinman, Structural and functional characterization of programmed ribosomal frameshift signals in West Nile virus strains reveals high structural plasticity among cis-acting RNA elements. *J. Biol. Chem.* **291**, 15788–15795 (2016).
36. S. Yan, Q. Zhu, S. Jain, T. Schlick, Length-dependent motions of SARS-CoV-2 frameshifting RNA pseudoknot and alternative conformations suggest avenues for frameshifting suppression. *Nat. Commun.* **13**, 4284 (2022).
37. S. Jain, C. S. Bayrak, L. Petingi, T. Schlick, Dual graph partitioning highlights a small group of pseudoknot-containing RNA submotifs. *Genes* **9**, E371 (2018).
38. S. Jain, S. Saju, L. Petingi, T. Schlick, An extended dual graph library and partitioning algorithm applicable to pseudoknotted RNA structures. *Methods* **162–163**, 74–84 (2019).
39. L. Petingi, T. Schlick, Partitioning and classification of RNA secondary structures into pseudonotted and pseudoknot-free regions using a graph-theoretical approach. *IAENG Int. J. Comput. Sci.* **44**, 241–246 (2017).
40. S. Jain, Y. Tao, T. Schlick, Inverse folding with RNA-As-Graphs produces a large pool of candidate sequences with target topologies. *J. Struct. Biol.* **209**, 107438 (2020).
41. D. Singh, S. V. Yi, On the origin and evolution of SARS-CoV-2. *Exp. Mol. Med.* **53**, 537–547 (2021).
42. R. Dirks, N. Pierce, A partition function algorithm for nucleic acid secondary structure including pseudoknots. *J. Comput. Chem.* **24**, 1664–1677 (2003).
43. E. Rivas, J. Clements, S. R. Eddy, Estimating the power of sequence covariation for detecting conserved RNA structure. *Bioinformatics* **36**, 3072–3076 (2020).
44. S. Tavaré, Some probabilistic and statistical problems in the analysis of DNA sequences. *Some Math. Quest. Biol.: DNA Sequence Anal.* **17**, 57–86 (1986).
45. P. J. Bredenbeek *et al*., The primary structure and expression of the second open reading frame of the polymerase gene of the coronavirus MHV-A59: A highly conserved polymerase is expressed by an efficient ribosomal frameshifting mechanism. *Nucleic Acids Res.* **18**, 1825–1832 (1990).
46. I. Brierley, P. Digard, S. C. Inglis, Characterization of an efficient coronavirus ribosomal frameshifting signal: Requirement for an RNA pseudoknot. *Cell* **57**, 537–547 (1989).
47. J. Herald, S. Siddell, An "elaborated" pseudoknot is required for high frequency frameshifting during translation of HCV 229E polymerase mRNA. *Nucleic Acids Res.* **21**, 5838–5842 (1993).
48. R. Kocherhans, A. Bridgen, M. Ackermann, K. Tobler, Completion of the porcine epidemic diarrhoea coronavirus (PEDV) genome sequence. *Virus Genes* **23**, 137–144 (2001).
49. J. F. Eleouet *et al*., Complete sequence (20 kilobases) of the polyprotein-encoding gene 1 of transmissible gastroenteritis virus. *Virology* **206**, 817–822 (1995).
50. C. P. Cho, S. C. Lin, M. Y. Chou, H. T. Hsu, K. Y. Chang, Regulation of programmed ribosomal frameshifting by co-translational refolding RNA hairpins. *PLoS ONE* **8**, 1–11 (2013).
51. L. Pekarek *et al*., Cis-mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function. *Nucleic Acids Res.* **51**, 728–743 (2023).
52. M. S. Angela, M. S. Costello, A. H. Kettring, R. J. Wingo, Ribosome collisions alter frameshifting at translational reprogramming motifs in bacterial mRNAs. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 21769–21779 (2019).
53. H. Mei, S. Kosakovsky Pond, A. Nekrutenko, Stepwise evolution and exceptional conservation of ORF1A/B overlap in coronaviruses. *Mol. Biol. Evol.* **38**, 5678–5684 (2021).
54. J. N. Zadeh *et al*., NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32**, 170–173 (2011).
55. R. Pagès, P. Aboyoun, R. Gentleman, S. DebRoy, Biostrings: Efficient manipulation of biological strings. R package version 2.64.0 (2022).
56. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002).
57. M. Clamp, J. Cuff, S. Searle, G. Barton, The Jalview Java alignment editor. *Bioinformatics* **20**, 426–427 (2004).
58. A. Stamatakis, Raxml version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
59. D. Huson, C. Scornavacca, Dendroscope 3: An interactive tool for rooted phylogenetic trees and networks. *Syst. Biol.* **61**, 1061–1067 (2012).
60. S. Jain, A. Laederach, S. B. V. Ramos, T. Schlick, A pipeline for computational design of novel RNA-like topologies. *Nucleic Acids Res.* **46**, 7040–7051 (2018).
61. E. P. Nawrocki, S. R. Eddy, Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).