



Heterogeneous and multiple conformational transition pathways between pseudoknots of the SARS-CoV-2 frameshift element

Shuting Yan^a and Tamar Schlick^{a,b,c,d,1}

Affiliations are included on p. 10.

Edited by Victoria D'Souza, Harvard University, Cambridge, MA; received August 27, 2024; accepted November 25, 2024 by Editorial Board Member Michael F. Summers

Frameshifting is an essential mechanism employed by many viruses including coronaviruses to produce viral proteins from a compact RNA genome. It is facilitated by specific RNA folds in the frameshift element (FSE), which has emerged as an important therapeutic target. For SARS-CoV-2, a specific 3-stem pseudoknot has been identified to stimulate frameshifting. However, prior studies and our RNA-As-Graphs analysis coupled to chemical reactivity experiments revealed other folds, including a different pseudoknot. Although structural plasticity has been proposed to play a key role in frameshifting, paths between different FSE RNA folds have not been yet identified. Here, we capture atomic-level transition pathways between two key FSE pseudoknots by transition path sampling coupled to Markov State Modeling and our BOLAS free energy method. We reveal multiple transition paths within a heterogeneous, multihub conformational landscape. A shared folding mechanism involves RNA stem unpairing followed by a 5'-chain end release. Significantly, this pseudoknot transition critically tunes the tension through the RNA spacer region and places the viral RNA in the narrow ribosomal channel. Our work further explains the role of the alternative pseudoknot in ribosomal pausing and clarifies why the experimentally captured pseudoknot is preferred for frameshifting. Our capturing of this large-scale transition of RNA secondary and tertiary structure highlights the complex pathways of biomolecules and the inherent multifarious aspects that viruses developed to ensure virulence and survival. This enhanced understanding of viral frameshifting also provides insights to target key transitions for therapeutic applications. Our methods are generally applicable to other large-scale biomolecular transitions.

Molecular dynamics | Conformational transitions | Transition path sampling | Programmed ribosomal frameshifting | Multiple RNA folds

Programmed ribosomal frameshifting (PRF) is an essential regulation mechanism of RNA transcript translation in viruses, also relevant to human genes (1–3). Coronaviruses employ PRF to direct the synthesis of their replicase proteins. Frameshifting is facilitated by specific RNA folds in the frameshifting element (FSE), which have emerged as therapeutic targets. For example, for SARS-CoV-2, a 3-stem RNA pseudoknot is believed to direct the elongating ribosome to pause and slip backward by one base over the heptameric “slippery site.” This backward shift of the mRNA transcript produces a new reading frame for protein synthesis. Though the 3-stem pseudoknot has been reported by various experimental and computational methods (4–12), especially at short sequences, alternative unknotted and knotted structures have been suggested for long sequence contexts (6, 13–15).

Why does a specific FSE fold facilitate frameshifting, and what is the role of alternative RNA folds? In prior work, we used our graph theory-based framework RNA-As-Graphs (RAG) (16–20) for RNA secondary structures to characterize different RNA FSE topologies. Our dual graphs represent RNA stems and helices as graph vertices and loops as edges. In our RNA dual graph library, RNAs with 3-stems are denoted using graph IDs, for example, 3_3, 3_5, and 3_6, to represent topologies/motifs of increasing levels of compactness. Besides the experimentally confirmed pseudoknot (3_6 dual graph), we identified, with support of chemical probing data, an alternative pseudoknot 3_3 (Fig. 1) involved in the transition targeted here, and a 3-way junction 3_5 (6, 7). The three possible folds share Stems 1 and 3, while the flexible Stem 2 is involved in a switch among them (6). Our atomic molecular dynamics (MD) simulations of all three folds at different RNA lengths revealed fluctuations from L-shaped to linear-shaped pseudoknots for 3_6 with hydrogen-bonding stabilizing networks (21), consistent with cryo-EM and crystallography studies (4, 5, 8, 9) and length-dependent

Significance

Frameshifting is an essential viral replication mechanism for compact RNA genomes. For SARS-CoV-2, a specific 3-stem RNA pseudoknot has been identified to stimulate frameshifting, but other folds have been suggested. Here, we capture atomic-level transition pathways between two key pseudoknots and demonstrate how they critically direct the placement of the viral RNA in the narrow ribosomal channel. Our work explains the role of the alternative pseudoknot in ribosomal pausing, clarifies why the experimentally captured pseudoknot is preferred for frameshifting, and provides insights to target key transitions for therapeutic applications. The capturing of a very large-scale RNA structure transition highlights complex biomolecular pathways and enhances our understanding of viral frameshifting.

Author contributions: S.Y. and T.S. designed research; S.Y. performed research; S.Y. contributed new reagents/analytic tools; S.Y. and T.S. analyzed data; and S.Y. and T.S. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission. V.D. is a guest editor invited by the Editorial Board.

Copyright © 2025 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

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

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2417479122/-DCSupplemental>.

Published January 24, 2025.

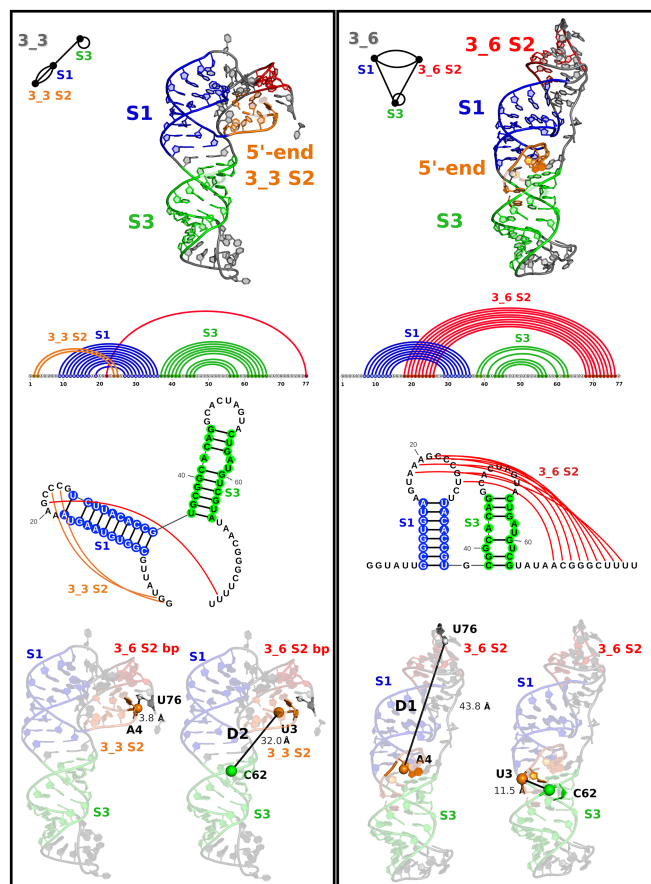


Fig. 1. The two 3-stem SARS-CoV-2 77-nt FSE pseudoknot motifs used in this work for simulating the transition. The target structure is dual graph: 3_6 (three helices with pseudoknot involving internal loop residues), and the starting structure is dual graph: 3_3 (two intertwined helices with an addition stem-loop). Dual graphs, 2D structures, and arc plots are shown. The distance variables D1 and D2 (order parameters) that characterize the two states are shown for each pseudoknot with related atoms as spheres: distance between O4' in A4 and U76 nucleobase center of mass, and distance between P atom in U3 and P atom in C62. Stems are labeled (S1 for Stem 1, 3_3 S2 or 3_6 S2 for Stem 2 in 3_3 or 3_6, and S3 for Stem 3).

stem interactions that suggest FSE transition pathways during ribosomal translation (21).

The two pseudoknots dominate at different sequence lengths (6, 15, 22). Up to 77-nt, the FSE typically adopts a 3_6 pseudoknot (8, 9, 11, 12, 23, 24), but once the slippery site and few upstream residues are appended, the alternative pseudoknot 3_3 dominates (6). Incorporating more upstream and downstream residues causes the central FSE region to revert to a 3_6 topology (6, 15, 22), observed in other beta coronaviruses, Sarbe subgenera, in our evolutionary analysis (22). Yet when the upstream sequences extend beyond 30-nt, an Alternative Stem 1 disrupts the pseudoknots and favors simple stem-loop folds (15), in agreement with chemical probing for long sequences (6, 7, 13).

The relevance of transitions to the frameshifting mechanism is evident from nearly abolished frameshifting for our transition-suppressing mutants (25). Namely, our structural and functional measurements for designed mutants that stabilize the 3_6, 3_3, and 3_5 FSE folds confirm single conformational states and show an order of magnitude reduction in frameshifting efficiency, underscoring the connection between transitions and frameshifting mechanism.

Besides these structural transitions, how the FSE fold interacts with the ribosomal machinery to pause over the slippery site remains unclear (26–29). The tension from ribosomal unwinding, critically affected by the length of the spacer region at the 5'-end of FSE, may be a factor (29, 30). Ribosome mediated folding and refolding between different FSE topological motifs, and the RNA's structural plasticity (31–33), appear crucial to frameshifting. Indeed, our designed transition suppressing mutants (6, 23) decrease frameshifting by an order of magnitude (7, 25). Thus, unraveling the FSE conformational and topological transitions is crucial for identifying the factors that influence frameshifting.

Here, we simulate at atomic detail the transition paths from 3_3 to 3_6 pseudoknots using transition path sampling (TPS), a method developed by Chandler and coworkers (34, 35) to capture rare events in the free energy landscape. Essentially, a Metropolis Monte Carlo procedure in MD trajectory space is performed, and many short unbiased MD trajectories are pieced together to cover the reaction. TPS does not alter the true potential energy surface, nor the underlying physical dynamics. However, detecting and correctly assembling all biochemically relevant transition regions in a complex RNA pseudoknot system using TPS poses a significant challenge, necessitating the development of new procedures for effective sampling coverage and free-energy computations. In the first application of TPS to a biomolecular system, a DNA polymerase β repair system (36) as highlighted in ref. 37, we defined a series of order parameters for multiple windows, each of which captures a local transition involving an angle, distance, or related geometric quantity, informed by prior MD simulations (38–40). Here, we divide the entire transition into overlapped windows defined by two distance order parameters (Fig. 1). These parameters capture the motions of the RNA 5'- and 3'-strand ends, which distinguish the 3_3 and 3_6 pseudoknots. Importantly, as starting structure, we use the sequence of our 3_6 “pseudoknot strengthening mutant” we designed using inverse folding (6); this mutant was confirmed experimentally to fold onto this pseudoknot without alternative folds (6, 25). After we cover the complete path by piecing together the overlapping windows, we use Markov State Modeling (MSM) combined with conformational clustering to describe the multiple pathways between conformational states in 5 families. These applications require tailored strategies for the overlapped windows and large-scale transition (*Materials and Methods*). Our free energy method BOLAS (36) is then used to calculate efficiently the free energy landscape with the same TPS order parameters.

Our path sampling of the SARS-CoV-2 FSE transition captures the initial Stem 2 unpairing by move of the 5'-end from global helix bending, triggered by the spacer region tension; base stacking and hydrogen bonding follows to create a new stem intertwining. Significantly, the calculated transition profile reveals a network of various conformations rather than a single sequential conformational rearrangement. The breaking of the pseudoknot stem, and rearrangement of stems and the dangling 5'-end, position the pseudoknot well in the narrow ribosomal channel, explaining why this pseudoknot is favored. Our characterization of the complex system by a Markov State Model and conformational families reveals an intricate transition network with major hubs of transition that define a system of checks and balances within the FSE conformational landscape, influencing frameshifting. Besides explaining how the spacer region tension triggers the transition which plays a role in frameshifting, our work has implications to design of conditions

that affect the spacer region, generated tension, and hence frameshifting efficiency. The combination of tailored methods are also generally applicable to other complex biomolecular systems.

Results

To explore how the FSE pseudoknot might interconvert from one pseudoknot to another during frameshifting (21), we study the 3₃ to 3₆ pseudoknot transition via TPS. Starting from our pseudoknot strengthening mutant sequence, designed to fold onto the 3₆ pseudoknot, as confirmed by SHAPE (6) and DMS-MaPseq (25) experiments, we also used insights from our MD (21) to set up an initial 3D model close to an intermediate 3₃ conformation. This initial conformation contains 3₃ Stem 2 base pairing as well as 3₆ S2 interactions. We generate initial points via Targeted MD (TMD) and perform TPS simulations on overlapped moving windows (shown in *SI Appendix*, Fig. S1). We track the transition by movement of the 5'-end using two distance variables: a A4-U76 distance that is unique to the 3₃ intermediate due to the triplex interactions, and a U3-C62 distance that describes the 5'-end extension along S3 helix in 3₆. In each window, we generate an ensemble of paths connecting predefined states using the two distance order parameters (*Materials and Methods* and *SI Appendix*).

Heterogeneous FSE Conformational Landscape. From successful trajectory series, we recover a heterogeneous conformational landscape connecting the 3₃ to 3₆ FSE pseudoknot transition (Figs. 2–4). All trajectories share a 5'-end movement accompanied by stem junction opening/closing, and 3₆ S2 base pairing rearrangement. To describe the complex dynamics, we apply time-lagged independent component analysis (TICA) and K-means clustering to define common metastates. We further classify our resulting 75 FSE clusters into five large conformational families in Fig. 2. These groups include (1) 3₃ triplex structures similar to the starting conformation, (2) 3₆ triplex structures after the breaking of 3₃ S2, (3) 3₆ linear structures with 5'-end at the stem junction, (4) 3₆ L-shape structures with 3₆ S2 bending, and (5) 3₆ C-shape highly compact structures (minor group). The “triplex” structures are grouped because they contain triplex interactions shared by Stem 2 of 3₃ and Stem 2 of 3₆. The linear 3₆ contains structures with the 5'-end dissociated from 3₃ S2 and located in the stem junction, representing the transition state between a stable 3₆ and 3₃. The C-shape 3₆ is highly stable. The heterogeneous and complex conformational free-energy landscape defines multiple transition paths.

Multiple Transition Pathways Revealed by Markov State Modeling. To define the transition paths from our complex ensemble, we build a Markov State Model from the defined clusters. The network graph in Fig. 3 is built from the transition matrix of connected states. Various transition paths are shown as produced by MSM by connecting trajectories that start from one of the 3₃ starting structures and end with one of the 3₆ end structures. Four representative transition paths analyzed in this work are highlighted and shown as series of arcs in Fig. 3 (Pink: 23 → 55 → 35 → 39; Green: 18 → 55 → 71 → 73 → 11 → 49; Purple: 23 → 55 → 6 → 40 → 71 → 26; and Yellow: 18 → 55 → 71 → 47).

The selected 3₃ and 3₆ structures are located at the outer ring of the transition network map in Fig. 3. At center, we identify

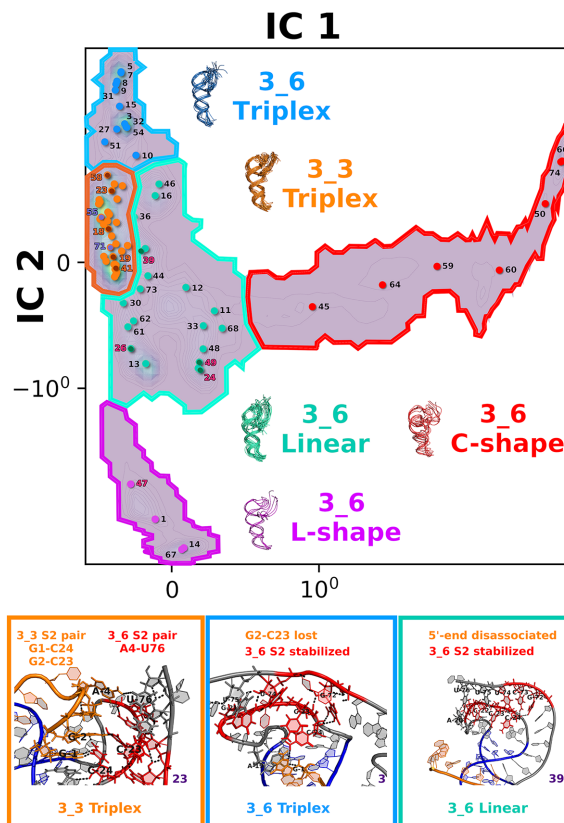


Fig. 2. Clusters and associated representative structures from the TPS harvested trajectories projected on the first two TICA independent components (ICs). The sampling projected on two ICs falls into 5 groups in terms of topology and global shape, including 3₃ triplex (orange), 3₆ triplex (blue), linear 3₆ (cyan), L-shape 3₆ (purple) and C-shape 3₆ (red). Screenshots of selected representative structures in each group are shown.

intermediates visited frequently by multiple transition paths. Key clusters 55 and 71 form when 3₃ S2 base pairs break but the 5'-end has yet to move further apart. The selected 3₃ structures are interconvertible, while selected 3₆ structures differ by the dangling 5'-end. We also map four representative paths in Fig. 3 as Green: 5'-end extended journey path, Pink: 5'-end rotation driven path, Yellow: 3'-end stretching driven path, and Purple: 3'-end driven bending path (*SI Appendix*, Text).

Central hubs in the network. In the transition network of Fig. 3, Clusters 55 and 71 are the main hubs that bridge 3₃ or 3₆-like conformations and connect heterogeneous transition paths.

Cluster 55 contains one base pair of 3₃ Stem 2 G1-C24, where G1 is π stacked with G25, thus keeping G1 in place and allowing G2 to move and pair with C23 for 3₃ S2. When the G1 interactions with C24 and G25 are broken, 3₃ S2 is eliminated and the transition to 3₆ is triggered.

Similarly, in Cluster 71, where G1 from 3₃ Stem 2 is near A19 and farther from the 3₃ S2 3'-strand residues (residues 21 to 24), H-bonds between G1 and U28 stabilize G1 at the stem junction instead of the competing S2 region. This altered orientation of 5'-end does not support a 3₃ S2 re-formation.

Two main changes characterize the transition: breaking of 3₃ S2, and rearrangement of the 5'-end.

Breaking of 3₃ S2 in various paths. The 3₃ pseudoknot Stem 2 break can occur by stem opening or shifting, both of which alter the 5'-end bases. Screenshots in Fig. 4 illustrate how the 3₃ S2 break triggers 3'-end stretching (Pink Path), Loop 1 relaxation (Green Path), and global stretching (Yellow Path). For

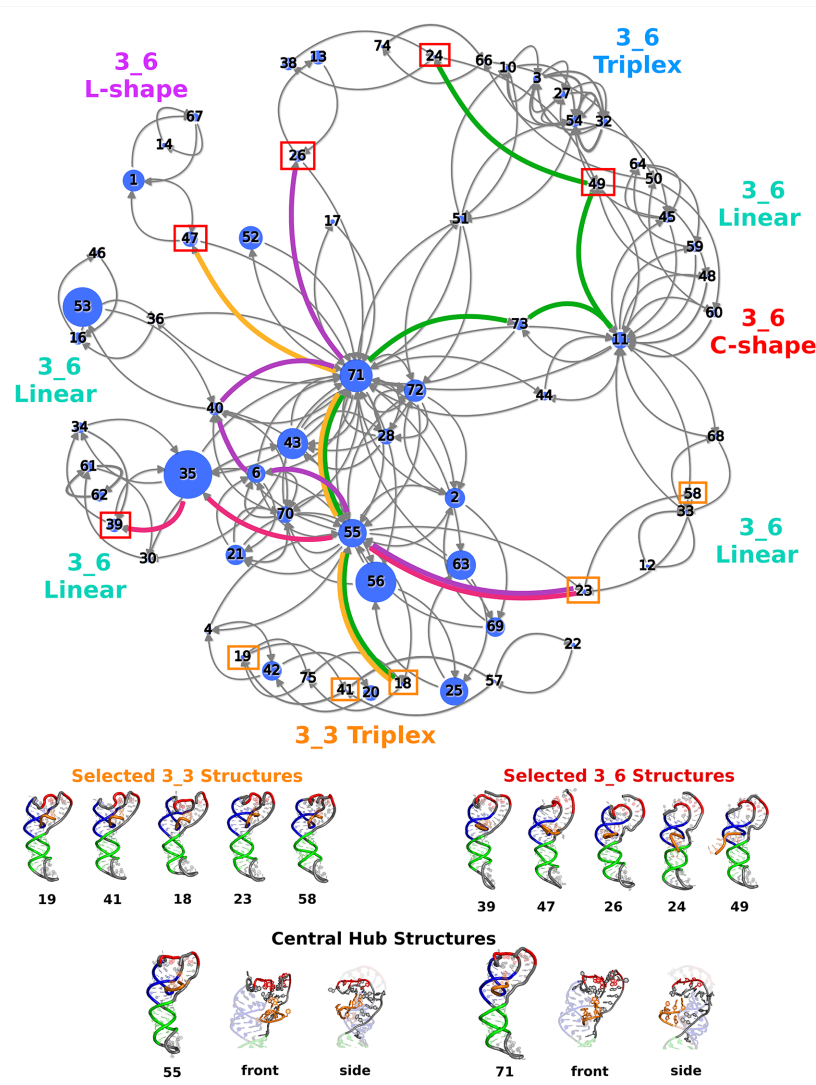


Fig. 3. Transition map from Markov State Modeling describing the 3₃ to 3₆ transition via multiple routes through different hubs and clusters of conformations. The connected states are shown as blue nodes, with areas proportional to the stationary probability of each state. Screenshots of selected starting and end structures (outer nodes) are shown at *Bottom*: 3₃ triplex in yellow (yellow rectangles in the transition map) and various 3₆ structures in red (red squares). The five 3₃ clusters have the lowest RMSD to the initial 3₃ iFold structure and 3₃ S2 base pairs. The five 3₆ clusters have low RMSD to the target 3₆ 3D model and 3₆ S2 base pairs. In addition, screenshots of central hub clusters 55 and 71 are shown, with front and side views of the enlarged pseudoknot stem region.

the Pink Path, the breaking of H-bonds between G2/U3 and A19 frees the 5'-end observed in a 2-ns trajectory (see [Movie S6](#)). The two mechanisms may occur at different ribosomal situations: Stem breaking might occur by ribosome fluctuations when FSE/ribosome interactions are tight, whereas base pair shifting might occur when the FSE is at the entry of the mRNA channel with unwound spacer and slippery region.

Relocation of the 5'-end in different 3₆ variations. Following the stem 2 break, the 5'-strand end residue motions play an important role in defining the junction ring and the global change in RNA shape. The interactions of 5'-end residues with U66 and A67 at the stem junction affect the conformation of the stem junction ring and the global shape (Fig. 4). In the four paths (*SI Appendix*), the 5'-end can cause local or global change to the FSE. When the 5'-end rotates and interacts with bases at the stem junction, the junction ring opens and closes along with the switch of FSE between linear and L-shape.

In the Green Path, the folded 5'-end moves further apart from S2, near U66 at S1/S3 junction. With the relaxation and rotation

of 5'-end, the junction ring closes up (A67 moves very close to S1 residues; see [Movie S7](#)).

In the Yellow Path, the folded 5'-end is near U66, similar to the green path. However, A67 is flipped out, leaving the stem junction still open.

In the Purple Path, when the folded 5'-end at S1/S3 junction interacts with A67 through stacking (G1-A67), the 5'-end remains folded and its movement expands the S1/S3 junction groove and causes the global conformational change from linear to L-shape (see [Movie S8](#)).

Free Energy Landscape of the Transition Path. We approximate the free energy landscape by the BOLAS algorithm of Radhakrishnan and Schlick (36), which combines the idea of TPS with umbrella sampling to estimate free energies with high accuracy. BOLAS considers trajectories that connect the start and end states, as well as other trajectories that visit the predefined transition region (*Materials and Methods*). Fig. 5 shows that the free energy path is path dependent but that surpassing central

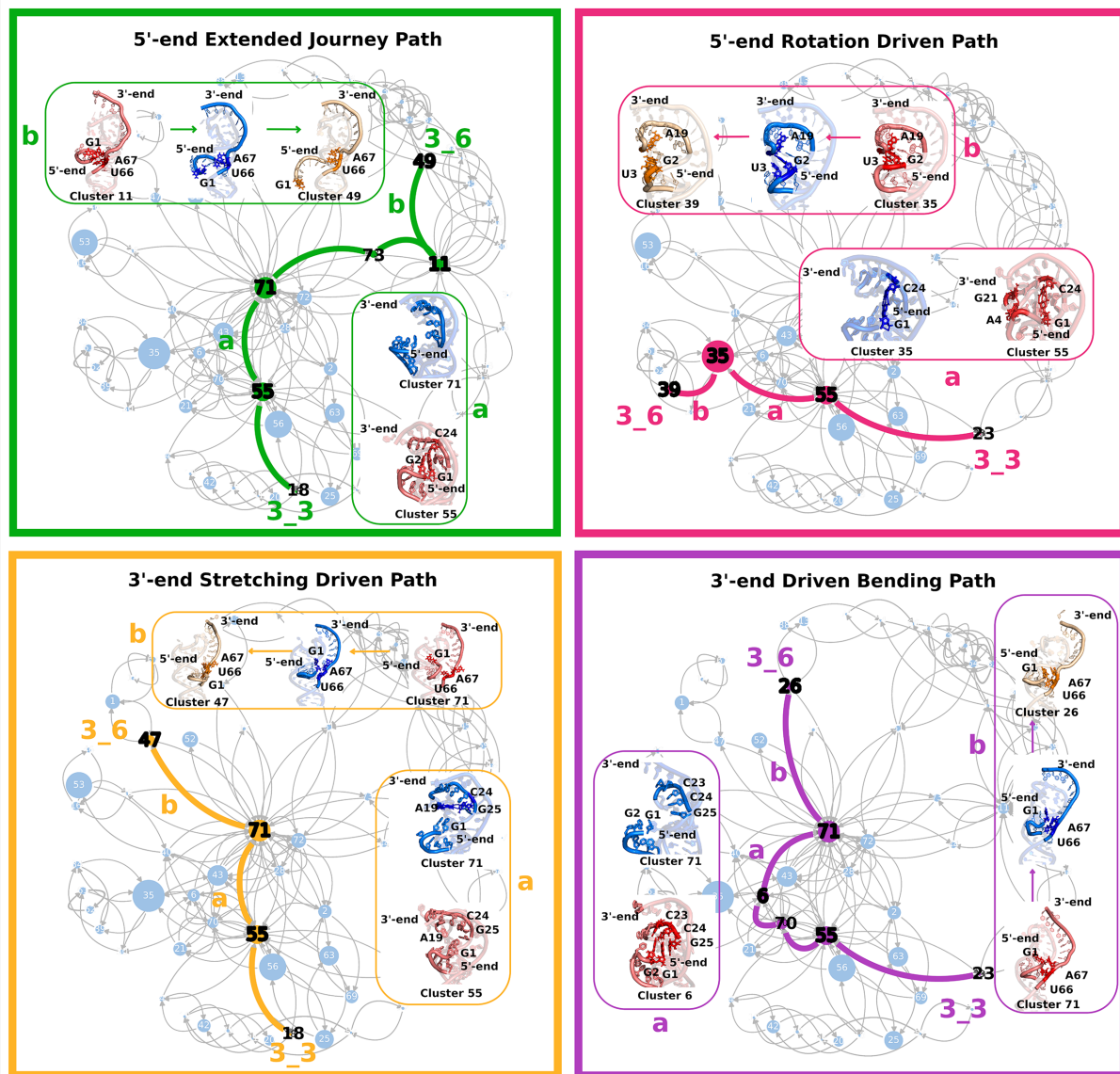


Fig. 4. Screenshots of the triplex and stem junction regions, including the 3_3 S2 breaking and the 5'-end relocation, of selected structures in our representative four paths. Transitioning from the red to blue conformations disrupts 3_3 S2 (labeled as a). Essential base pairs are labeled, and hydrogen bonds (H-bonds) are depicted using dashed lines. Transitioning from the red to blue to yellow conformation in each path, the 5'-end explores various positions (labeled as b). Key residues are highlighted.

hubs 55 and 71 is associated with a high barrier. The free energy landscapes also show that breaking 3_3 S2 (the first two basins at top left) tends to be more difficult than the structural relaxation and reorganization of 3_6 (including stabilizing 3_6 S2 and movement of 5'-end).

The free energy trajectories for our four selected paths in Fig. 5, bottom, highlight triplex and stem junction regions; see enlarged screenshots of key structures in Fig. 4.

Among the four selected paths, the Green Path exhibits more significant conformational changes as reflected by the free energy evolution (Figs. 4 and 5). These include the breaking of 3_3 S2 (increase of free energy at Frame 1,925), relaxation of Loop 1, rearrangement of 3_6 S2, and opening and closing of stem junction ring (last three largest barriers). In contrast, while the 5'-end freely explores the space after 3_3 S2 breaking in the other three paths, G7 and U66 remain stacked (see detailed description of four paths in *SI Appendix*). However, this stacking is lost in the Green Path, allowing for the emergence of other potential

conformational states. Thus, the stacking disruption between the 5'-end residues and stem junction bases is likely required for a more stable 3_6 conformation to form. See *SI Appendix* for discussions of other paths.

Proposed Mechanism of the RNA Transitions during Frameshifting. In Fig. 6, the 3_3 starting and 3_6 end structures are aligned and superimposed with the Cryo-EM structure of a translating ribosome primed for frameshifting [PDB: 7O7Z (4)]; the consequences of our transition on the spacer and ribosome channel placement are analyzed.

To compare the orientations of our 3_3 start and 3_6 end structures with the FSE solved in 7O7Z, we zoom in to focus on pseudoknot stem junction and the 5'-end.

The superimposition of our 3_3 and 3_6 pseudoknots with the Cryo-EM FSE reveals clashes with ribosomal proteins, where the 3_3 FSE shows much greater clashes with proteins S3 and eS10. In 3_3 S1, the base pair U36-A65 stacks with G35-C8,

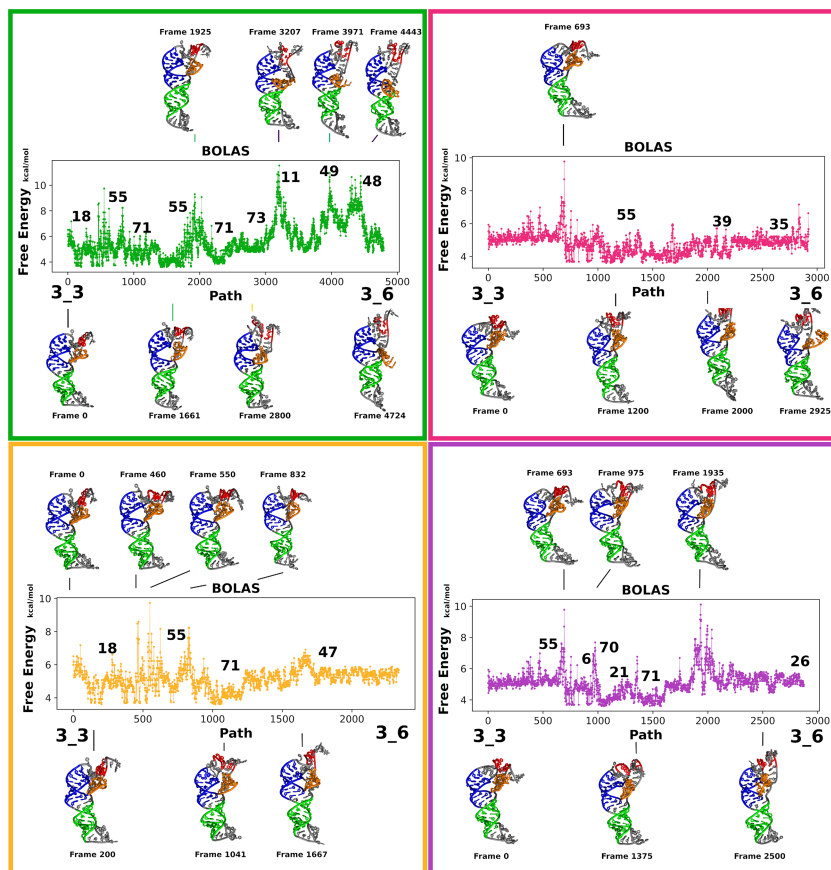
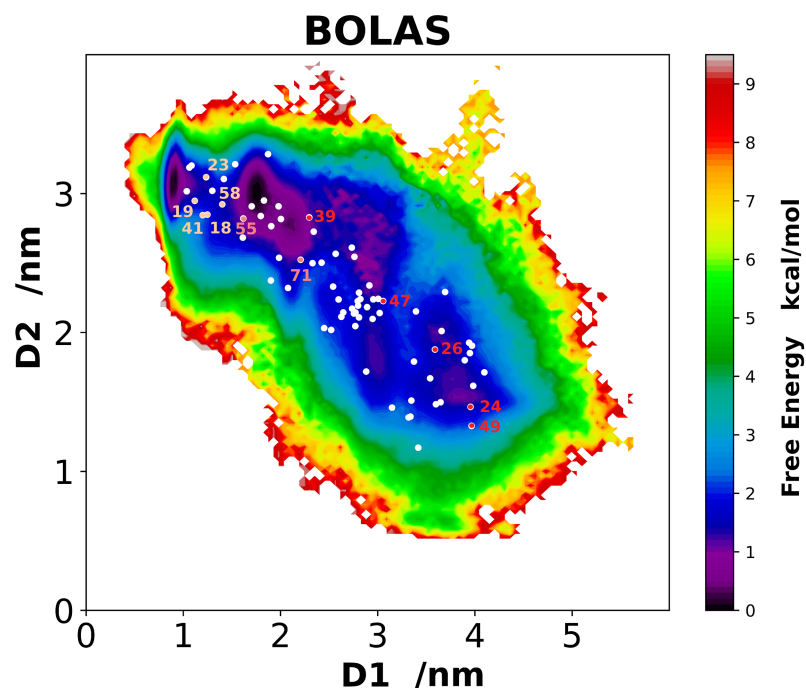


Fig. 5. Free energy landscapes of 3_3 to 3_6 transition in the 2D order parameter space. (*Top*) Selected 3_3 and 3_6 clusters, as well as the two center hubs, are labeled. The 3_3 clusters are colored in orange, 3_6 clusters are colored in red and the two central hubs are colored in salmon. (*Bottom*) 1D free energy evolution and screenshots of key structures in the transition along each path.

while it is broken in linear 3_6 and forms a part of S3 in the L-shaped 3_6 as in the Cryo-EM structure. This rearrangement is a consequence of the 5'-end rotation due to pulling and relocation

in the mRNA channel. This tight fit might also explain why the frameshifting efficiency is correlated to the stability of the first few base pairs in the helix stem of FSE pseudoknot.

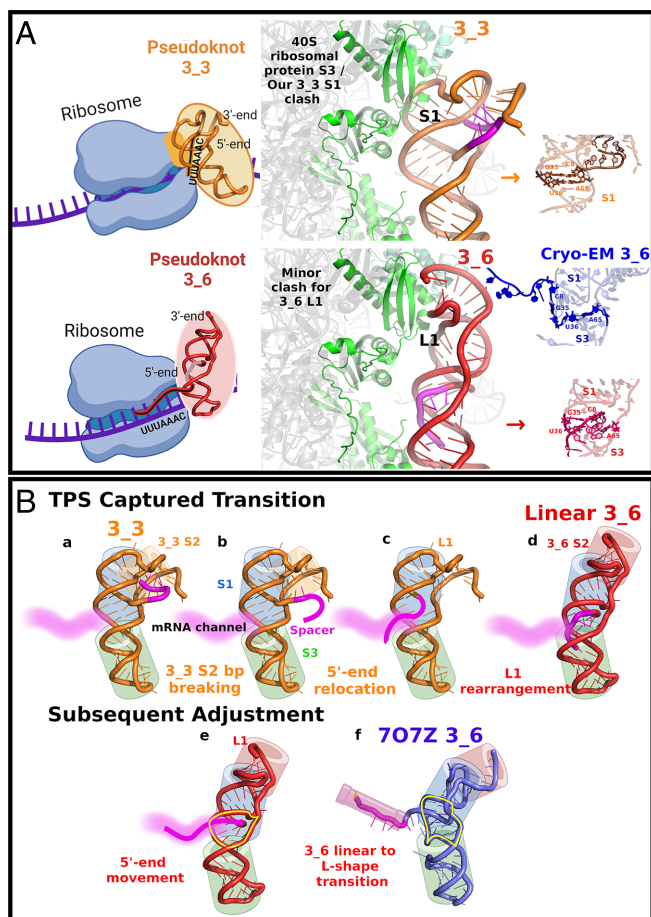


Fig. 6. The 3₃ starting and 3₆ end structures aligned and superimposed with Cryo-EM structure of a translating ribosome primed for frameshifting [PDB: 707Z (4)] and consequences of transition on spacer and ribosome channel placement. (A) We superimpose our 3₃ start structure (orange) and our 3₆ end structure (red) with the FSE in 707Z (blue). The small ribosome subunit (SSU) is colored in gray. The FSE pseudoknot (L-shaped 3₆) in 707Z is colored in blue. Green-colored ribosomal proteins, including uS3, uS5, eS30, and eS10, interact with the FSE, as illustrated by Bhatt et al. (4). At right, we zoom in on the S1/S3 junction and the 5'-end. (B) The proposed mechanism from 3₃ to linear 3₆ and eventually L-shaped 3₆.

Hence, we propose the following mechanism for transition from 3₃ to linear 3₆ and eventually L-shaped 3₆: The spacer at the 5'-end is released from 3₃ S2 and moves to the 3-stem junction; when straightened, it fits into the mRNA channel well in our 3₆ pseudoknot end state. Further pulling from translation triggers the transition to an L-shaped 3₆, as seen in the Cryo-EM structure. This L-shaped 3₆, along with breaking of some junction base pairs, organizes the narrow mRNA channel optimally.

Discussion

Our successful TPS application on a very complex and significant RNA secondary structure conformational change, requiring many GPUs and tens of thousands of trajectories, in combination with conformational clustering, Markov State Modeling, and BOLAS free energy computations, has allowed us to shed light on the complex kinetics underlying RNA secondary and tertiary topology changes at atomic resolution. The uncovered intermediate states and kinetic bottlenecks offer a deeper understanding of the folding landscapes of the SARS-CoV-2 FSE. The resulting

insights into viral frameshifting contribute to our understanding of the factors that may influence pausing.

The multiple possible 3₆ conformations and transition paths from a 3₃ to 3₆ pseudoknot are important because they involve base pairing of the 5'-end, which is the key spacing region between the slippery site and downstream stems. The spacer determines the positioning of the slippery sequence and the tension from translocation on the mRNA during frameshifting. The length of the spacer determines the correct positioning of FSE for optimal frameshifting efficiency when the ribosome is on the slippery site (41). Changes in the spacer length alter the base pair during the initial encounter of the ribosome with the slippery site (42) and may require the FSE to be present at a specific time of translation (43). This “effective spacer length” is shortened when the 3₃ pseudoknot stem forms, as we observe in Fig. 6 by aligning our selected FSE conformations along transition pathways with that of the ribosome complex Cryo-EM structure (4). A shorter spacer increases the tension on the mRNA (41). When the FSE is not appropriately unwound during the elongation cycle, frame maintenance could potentially be compromised through the generation of tension in the mRNA, effectively pulling the mRNA in a 3' direction while promoting breakage of the tRNA anticodon:codon interaction and realignment of the tRNA in the 5' direction (41).

Hence, we propose that the formation of 3₃ aids in tuning the tension through the spacer region. Specifically, the orientation and conformation of the spacer impact the surrounding residues at the three-stem junction and can enhance or release the generated tension during frameshifting. When the spacer is folded at the junction, as in the 3₃ conformation, tension is increased, causing global shape change; when global or 3'-strand stretching leads to a stabilized 3₆ pseudoknot, tension is released (see Green Path and *SI Appendix, Figs. S1 and S5* from transitions observed in a TPS window).

Importantly, this tension and the FSE fold also alter critically how the FSE interacts with the ribosome (Fig. 6). In the 3₃ fold, the 5'-end clashes with the narrow mRNA channel, whereas the 3₆ fold fits there well (Fig. 6A). Indeed, because the mRNA tunnel has an average diameter smaller than that of an RNA double helix (44), the folded spacer in the intermediate conformations we capture must be pulled straight, followed by the breaking of 3₃ S2. A contraction of the mRNA entry channel during translocation (45) might correspond to the 3₃ to 3₆ transition. In Fig. 6, we see that 3₃ S2 clashes with ribosomal proteins at the mRNA channel entry, so pulling the 5'-end would place the slippery site and spacer within the narrow mRNA channel as single strands. Relief of this clash triggers the 3₃ pseudoknot to linear 3₆ pseudoknot transition, and further to the L-shaped 3₆ pseudoknot captured in the complex; U36-A65 base pair in S1 breaks and reforms as U36 rotates due to the 5'-end pulling.

Our transition paths also explain why the 3₆ pseudoknot is favored over 3₃ for enhancing frameshifting and inducing translocation pause. Because the first stem near the entrance of ribosomal mRNA channel is S1 in 3₆ but S2 in 3₃, the latter is easier to break because of the free 5'-end in 3₃ versus the threaded 5'-end in 3₆. Further, because ribosomal pausing depends on the FSE fold, specifically the number of base pairs that must break (43), the 3₆ pseudoknot may be preferred. Further, we have shown (*SI Appendix, Fig. S3*) that 3₃ S2 breaking can occur by strand opening as well as base pair shifting, as also suggested experimentally (44). Thus, during ribosomal mRNA unwinding, thermal fluctuations and strand pulling can more easily break S2 of 3₃. This is in contrast to breaking S2 of 3₆, which can only be achieved by strand pulling.

Besides mechanistic insights, which correlate well with and expand upon experimental observations, direct applications of our findings are potential mutant sites in residues that play an important role in the transitions. Residues U66 and A67 affect the conformation of stem junction ring through their interactions with the dangling 5'-end (Fig. 4). The dynamic interplay between the opening and closing of the stem junction is significant for therapeutic applications, as this region represents a potential target site for small molecule binding. Similarly, G1 and U77 are pivotal in stabilizing the triplex structure when incorporated within it. Thus, residues G1, U66, A67, U76, and U77 can be targeted to suppress transitions through mutations designed to interfere with these critical tertiary contacts between the 5'-end and 3'-end sugar backbones. Further work may also improve relating indirect kinetic measurements of ribosomal pausing to transition rates.

Certainly, given the complexity of our RNA system, and the numerous conformational transitions that exist on the secondary structure level, we cannot rule out smaller transition state regions, kinetic traps, or insufficient sampling in the low coverage windows. Indeed, the 3₃ to 3₆ pseudoknot transition we captured is one in a complex landscape where other transitions involving alternative RNA folds may be involved. Clearly, the distance variables that define our transition simplify our sampling and convergence assessment. General path variables (46, 47) can be considered in further FSE transitions and analyses. Nonetheless, our robust thermodynamically reliable pathways and general findings generated from thousands of independent short simulations that successfully connect and piece together the target endpoints are valuable. Our combined toolkit (transition path sampling, Markov State Modeling, and cluster analysis for multiple overlapped windows, BOLAS free energy computations) is generally applicable to other biomolecular processes. The advantage of BOLAS here is that it can use coordinate variables from TPS with lower errors than umbrella sampling (36).

In conclusion, our application of TPS and MSM to study the folding dynamics of a biologically important RNA structure, the frameshifting RNA element of the SARS-CoV-2 virus, has yielded invaluable insights into the intricate pathways and kinetic mechanisms between two FSE pseudoknots. By dissecting infrequent but biologically crucial transition events and exploring the energetic landscapes, our successful approaches of TPS and BOLAS have contributed to the understanding of the FSE structural plasticity and the impact of this topological transition on the spacer region and resulting tension required for proper fitting into the mRNA ribosomal channel, and in turn frameshifting efficiency. Further structural and dynamic integration with experimental data on ribosomal pausing and frameshifting efficiency can help unravel the intricate frameshifting dynamics in viral RNA elements and explore our proposed avenues for therapeutic interventions.

Materials and Methods

Starting and Target Structures. From previous work, the 3₆ 3-stem pseudoknot is dominant for short sequences, such as 77-nt, while the 3₃ alternative pseudoknot dominates for long sequences with additional upstream and downstream residues. A possible mechanism was proposed (21) that involves translocating ribosome unwinding the base pairs involving 5'-end and breaking 3₃ S2, leading to 3₆ pseudoknot. Our proposed intermediate structure contains both 3₆ and 3₃ S2 interactions (21). Given the potential complex landscape, we investigate the transition between 3₃ and 3₆ pseudoknot motifs from this potential intermediate structure to 3₆. For a representative stable 3₆ target, we used the pseudoknot strengthening motif

(PSM) designed in our previous work (6) via inverse folding algorithm RAG-IF (48), which has been examined to fold onto the 3₆ 3-stem pseudoknot at 100% (6) and thus guarantee a funneled 3₆ landscape. The 3₆ PSM sequence include mutations [G3U, U4A, G18A, C19A, C68A, A69C] to favor 3₆ S2 against 3₅ S2. Its 3D structure contains a 9-base pair 3₆ S2 and the stability has been confirmed via MD simulations (21) [more 3D prediction details of the two 3D models are available in our previous publication (21)].

System Preparation and Initial Runs. The two 3D models are solvated and centered in a rectangular water box 10 from the edge, and the system is neutralized with K⁺ and additional ions for a concentration of 0.15M KCl. The resulting systems contain 92,179 atoms, including the frameshift RNA (2,465 atoms), 30,002 water molecules, 86 Cl⁻, and 162 K⁺. The intermediate 3₃ structure is the cluster center structure from previous microsecond MD simulation, and the 3₆ PSM system is equilibrated via Amber OL3 force field (49). It is unclear whether Mg²⁺ ions are essential for the SARS-CoV-2 FSE (10). Moreover, Mg²⁺ ions stabilize RNA folds, which might affect the observation of conformational transitions we are interested in. Thus, Mg²⁺ ions are not introduced in the systems as in our previous MD work (21).

To generate the initial path, targeted MD trajectories are conducted with Amber 22 (50–53), implemented with a constraint constant 0.01 on RNA atoms. With a decrease of RNA RMSD with reference to 3₆ PSM structure over 100 ps at 1 fs timestep, the conformational change is driven from the intermediate 3₃ structure to 3₆. Although unphysical, the targeted trajectory helps define starting points for TPS. From the targeted MD trajectory, we identify the distances that show a linear increase/decrease, as well as correspond to interactions in the initial (3₃) and target (3₆) states.

Performing Path Sampling. We perform TPS from initial paths as 10 ps sliding windows incrementally advanced along the 100 ps targeted MD trajectory by 2.5 ps at a time to ensure that a full path can be obtained from continuous transitions. We apply the shooting/shifting protocol that interfaces with Amber 22 (50) for each pair of unbiased simulations. Each pair of simulations are conducted under NPT (1 bar and 298.15 K) ensembles for 2 ns in total. Simulations are run with a timestep of 1 fs and a SHAKE algorithm (54) with constraints on bonds containing hydrogen.

Path segments corresponding to transitions in each window are defined in *SI Appendix, Table S1*. Trajectories in each window are harvested using the shooting algorithm to connect two metastable states A and B by means of a Monte Carlo protocol in trajectory space (35). In each window, TPS is performed for a maximum of 301 iterations with a preset desired acceptance rate of 0.33. Selected trajectory pairs of interest are extended to 100 ns each to examine whether the states being sampled are stable or experience continuing transitions.

All MD simulations were conducted on the Greene HPC clusters at the New York University High Performance Computing facilities. The TMD initial trajectories are performed on the CPU nodes on Greene, which are equipped with Lenovo SD650 model and contain 48 CPUs/node and 192 GB RAM. Each unbiased MD simulation is performed on the NVIDIA GPU nodes on Greene, which contain 4 GPUs on each node. Each MD simulation is run using one GPU with a performance of 90.92 ns/day. A total of 15,630 pairs of simulations (including both TPS and BOLAS runs) are performed for a cumulative computation time of 340 d.

Monitoring Adequate Sampling and Convergence. TPS runs in the windows with few harvested trajectories are extended to 601 iterations and replicated independently with a different random seed to ensure the transition region is well sampled. The convergence of harvested trajectories from TPS run in each window is examined by monitoring the L1 norm of 2D histogram from the sampling on the 2D order parameter space. The decrease of L1 norm of the collective 2D histogram and plateau before reaching 0 indicate the convergence of sampling from harvested trajectories.

Markov State Modeling. To extract the transition between different windows, we apply Markov State Modeling to investigate the transition network via PyEMMA (55). In the PyEMMA workflow, MD trajectories are processed and

discretized first. A Markov state model is estimated from the resulting discrete trajectories and validated. Detailed protocol includes the following steps:

Extracting molecular features from the raw data. The coordinates of key residues involved in the transition are selected variables, including the coordinates of phosphate atoms and centers of mass of residues in the 5'-end (residues 1 to 7), the L1 region (residues 18 to 37), and the 3'-end (residues 65 to 78) as well as pairwise distances between these phosphate atoms. The total number of variables is 975.

Transforming those features into a suitable, low dimensional subspace. Because of the large variety of conformational states, we had to use clustering to organize our results. Approaches other than clustering are possible. TICA is used to transform the features into a set of slow coordinates. TICA maximizes the autocorrelation of transformed coordinates and finds a maximally slow subspace, used as a dimension reduction technique from which the dominant components will be inputs for K-means clustering. To apply TICA (56, 57), the instantaneous ($\mathbf{C}(0)$) and time-lagged ($\mathbf{C}(\tau)$) covariance matrices with elements

$$c_{ij}(0) = \langle \tilde{x}_i(t) \tilde{x}_j(t) \rangle_t$$

$$c_{ij}(\tau) = \langle \tilde{x}_i(t) \tilde{x}_j(t + \tau) \rangle_t,$$

where $\tilde{x}_i(t)$ denotes the i th feature at time t after the mean has been removed.

After estimating the covariance matrices, TICA solves the generalized eigenvalue problem

$$\mathbf{C}(\tau)\mathbf{u}_i = \mathbf{C}(0)\lambda_i(\tau)\mathbf{u}_i,$$

where $i = 1, \dots, n$, to obtain independent component directions $\mathbf{u}_i$.

Dimension reduction is performed by projecting the (mean free) features $\tilde{\mathbf{x}}(t)$ onto the leading d independent components $\mathbf{U}_d = [\mathbf{u}_1 \dots \mathbf{u}_d]$,

$$\mathbf{y}(t) = \mathbf{U}_d^T \tilde{\mathbf{x}}(t),$$

where, in practice, d is preferably chosen such that a specific fraction of kinetic variance c_d is retained (e.g., 95%). In this work, the number of dimensions is reduced from 975 to 432 with lag time 0.2 ns.

Discretizing the low dimensional subsets into a state decomposition. Clustering is performed on the TICs to a set of states (75 clusters in this work) using K-means clustering algorithm.

Estimating a maximum likelihood or Bayesian MSM from the discrete trajectories and performing validation tests. Transitions are counted at the prespecified lag time (0.2 ns). Every pairwise transition at this lag time is counted and stored in a count matrix. Then, the count matrix is converted to a row-stochastic transition probability matrix $\mathbf{P}(\tau)$, which is defined for the specified lag time. For equilibrium MD simulations, $\mathbf{P}(\tau)$ should obey detailed balance which is enforced by constraining the estimation of $\mathbf{P}(\tau)$ to the following equations:

$$\pi_i p_{ij} = \pi_j p_{ji},$$

where π_i is the stationary probability of state i and p_{ij} is the probability of transitioning to state j conditional on being in state i .

The estimation is performed by plotting the implied timescales as a function of τ for selecting the MSM lag time, and the validation is performed via the Chapman-Kolmogorov test for checking whether a given transition probability matrix is approximately Markovian.

Once validated, the transition matrix can be decomposed into eigenvectors and eigenvalues. The highest eigenvalue, $\lambda_1(\tau)$, is unique and equal to 1. Its corresponding left eigenvector is the stationary distribution, π :

$$\pi^T \mathbf{P}(\tau) = \pi^T$$

The subsequent eigenvalues $\lambda_{i>1}(\tau)$ are real with absolute values less than 1 and are related to the characteristic or implied timescales of dynamical processes within the system. The dynamical processes themselves (for $i > 1$) are encoded by the right eigenvectors

$$\mathbf{P}(\tau)\psi_i = \lambda_i(\tau)\psi_i.$$

The coefficients of the eigenvectors represent the flux into and out of the Markov states that characterize the corresponding process.

The Bayesian MSM is estimated at lag time $\tau = 0.2$ ns in this work.

Analyzing the stationary and kinetic properties of the MSM. The mean first passage times and equilibrium probabilities on the metastable states are computed using PyEMMA (55). The network of fluxes from selected 3_3 states to selected 3_6 states, via a set of intermediate nodes, are computed, which can be used to compute transition pathways (and their weights) from one state (or a set of states A) to another state (or a set of states B). Fluxes can be computed in EMMA using transition path theory.

The Entire Transition Path. After obtaining the MSM transition network, we trace the transitions from 3_3 to 3_6 conformations by searching the shortest path and connecting trajectories that exhibit transitions among two or more clusters and meanwhile visit the same clusters. We select 5 clusters that have the smallest RMSD with reference of the 3_3 starting structure (clusters 23, 19, 18, 58, 41) as starting 3_3 conformations and 5 clusters that have 3_6 S2 base pairs and small RMSD from the 3_6 end structure (clusters 24, 26, 49, 47, 39) as the ending 3_6 conformations in the path search. Sequences of trajectories that start from any of the starting 3_3 conformations and end with any of the ending 3_6 conformations are accepted paths. These paths are further examined by checking the connectivity in the free energy values from (a) the MSM reweighted free energy surface, (b) direct TPS, and (c) BOLAS calculation, as well as sampling projection on the 2D order parameter space. Paths with discontinuation in the free energy values by any measurement or in the sampling projected on the 2D order parameter space are rejected, indicating the structures grouped in the same cluster from various trajectories exhibit large deviations and require further fine sampling for interpolation.

BOLAS Free Energy Calculation. To map the free energy landscape of the transition region, we employ the scheme "Biomolecular free energy profiles by a shooting/umbrella sampling protocol (BOLAS)" (36) that combines a Monte Carlo ensemble of trajectories generated by the shooting algorithm with umbrella sampling. "BOLAS" preserves microscopic reversibility and leads to the correct equilibrium distribution, resulting in free energy profiles along complex reaction coordinates for biomolecular systems with much higher accuracy than umbrella sampling. While the goal of TPS is to generate an ensemble of trajectories that connect A to B, a modified path action for "BOLAS" is required to compute the unbiased probability distributions of order parameters at equilibrium. In the TPS path action,

$$S\{\chi\}^\tau = \rho(0)h_A(\chi^0)H_B\{\chi\}^\tau,$$

where trajectories $\{\chi\}^\tau$ of length τ are generated by the Metropolis algorithm, the boundary conditions $h_A(\chi^0)$ and $H_B\{\chi\}^\tau$ introduce bias and prevent the correct estimation of the equilibrium probability distributions. This is because the following sampling cases will not be harvested by TPS algorithm: trajectories that visit A and not B; trajectories that visit B but not A; trajectories that visit neither A nor B. "BOLAS" aims to include these classes of trajectories with the following BOLAS path action:

$$S\{\chi\}^\tau = \rho(0).$$

To enhance the efficiency of computing the probability distribution $P(\chi_i)$ over a desired range of χ_i , the window-based umbrella sampling strategy is employed, where the BOLAS protocol is run independently in each of these windows. More detailed derivation of BOLAS can be found in the previous work (36).

BOLAS protocol works well with the overlapping window strategy used in our work to model RNA transitions. We apply it to obtain the free energy profile on the 2D order parameter space. Due to the complex RNA free energy landscape, the sampling becomes more easily trapped in certain states when running BOLAS calculations compared to TPS, for example structures tend to visit and get trapped in a linear 3_6 conformation, which is more similar to the end 3_6 target structure but further apart from the transition region. To focus on the transition region, we perform BOLAS protocol on the windows of TPS and adopt the same definitions of states A and B. The difference of a BOLAS run compared to TPS is that each BOLAS trajectory spans the predefined transition region (see [SI Appendix, Fig. S1](#)).

The combined methods, including TPS/BOLAS on overlapped windows and Markov State Modeling tailored and applied to a large-scale transition, are generally applicable to other biomolecular systems.

Data, Materials, and Software Availability. Code for TPS/BOLAS algorithm and scripts for MSM analysis have been deposited in GitHub Schlick Lab repository TPS-MSM (<https://github.com/Schlicklab/TPS-MSM>) (58).

ACKNOWLEDGMENTS. We thank Shereef Elmetwaly and Shenglong Wang for technical assistance and the New York University High-Performance Computing facilities and Simons Center for Computational Physical Chemistry for providing our group dedicated resources to perform this work. We gratefully acknowledge

funding from the NSF Rapid Response Research Award 2030377 from the Division of Mathematical Science and the Division of Chemistry, NSF Award DMS-2151777 from the Division of Mathematical Sciences, NSF CHE-2330628 from the Molecular Foundations for Biotechnology Program, NIH R35GM122562 Award from the National Institute of General Medical Sciences, and Philip-Morris International to T.S.

Author affiliations: ^aDepartment of Chemistry, New York University, New York, NY 10003; ^bDepartment of Mathematics and Computer Science, Courant Institute of Mathematical Sciences, New York University, New York, NY 10012; ^cNew York University - East China Normal University Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, People's Republic of China; and ^dSimons Center for Computational Physical Chemistry, New York University, New York, NY 10003

- J. D. Dinman, Programmed ribosomal frameshifting goes beyond viruses. *Microbe Wash DC* **1**, 521–527 (2006).
- N. Caliskan, F. Peske, M. V. Rodnina, Changed in translation: mRNA recoding by -1 programmed ribosomal frameshifting. *Trends Biochem. Sci.* **40**, 265–274 (2015).
- 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).
- P. R. Bhatt *et al.*, Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome. *Science* **372**, 1306–1313 (2021).
- 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).
- 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).
- 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).
- C. P. Jones, A. R. Ferré-D'Amaré, Crystal structure of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) frameshifting pseudoknot. *RNA* **28**, 239–249 (2022).
- 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).
- 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).
- A. Wacker *et al.*, Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy. *Nucleic Acids Res.* **48**, 12415–12435 (2020).
- 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).
- T. C. T. Lan *et al.*, Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells. *Nat. Commun.* **13**, 1128 (2022).
- 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).
- S. Lee, S. Yan, A. Dey, A. Laederach, T. Schlick, An intricate balancing act: Upstream and downstream frameshift co-regulatory elements. *Biochem.*, in press (2025).
- H. H. Gan, S. Pasquali, T. Schlick, Exploring the repertoire of RNA secondary motifs using graph theory: implications for RNA design. *Nucleic Acids Res.* **31**, 2926–2943 (2003).
- S. Pasquali, H. H. Gan, T. Schlick, Modular RNA architecture revealed by computational analysis of existing pseudoknots and ribosomal RNAs. *Nucleic Acids Res.* **33**, 1384–1398 (2005).
- J. Gevertz, H. H. Gan, T. Schlick, In vitro RNA random pools are not structurally diverse: A computational analysis. *RNA* **11**, 853–863 (2005).
- T. Schlick, Adventures with RNA graphs. *Methods* **143**, 16–33 (2018).
- G. Meng, M. Tariq, S. Jain, S. Elmetwaly, T. Schlick, RAG-Web: RNA structure prediction/design using RNA-As-Graphs. *Bioinformatics* **36**, 647–648 (2020).
- 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).
- S. Yan, Q. Zhu, J. Hohl, A. Dong, T. Schlick, Evolution of coronavirus frameshifting elements: Competing stem networks explain conservation and variability. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2221324120 (2023).
- 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).
- K. Neupane *et al.*, Structural dynamics of single SARS-CoV-2 pseudoknot molecules reveal topologically distinct conformers. *Nat. Commun.* **12**, 4749 (2021).
- A. Dey, S. Yan, T. Schlick, A. Laederach, Abolished frameshifting for predicted structure-stabilizing SARS-CoV-2 mutants: Implications to alternative conformations and their statistical structural analyses. *RNA* **30**, 1437–1450 (2024).
- T. Jacks, H. D. Madhani, F. R. Masiarz, H. E. Varmus, Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. *Cell* **55**, 447–458 (1988).
- J. A. Kelly, M. T. Woodside, J. D. Dinman, Programmed -1 ribosomal frameshifting in coronaviruses: A therapeutic target. *Virology* **554**, 75–82 (2021).
- E. P. Plant, J. D. Dinman, Torsional restraint: A new twist on frameshifting pseudoknots. *Nucleic Acids Res.* **33**, 1825–1833 (2005).
- E. P. Plant *et al.*, The 9-Å solution: How mRNA pseudoknots promote efficient programmed 1 ribosomal frameshifting. *RNA* **9**, 168–174 (2003).
- H. F. Noller, M. M. Yusupov, G. Z. Yusupova, A. Baucom, J. H. D. Cate, Translocation of tRNA during protein synthesis. *FEBS Lett.* **514**, 11–16 (2002).
- D. B. Ritchie, D. A. N. Foster, M. T. Woodside, Programmed -1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 16167–16172 (2012).
- 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).
- C. F. Hsu *et al.*, Formation of frameshift-stimulating RNA pseudoknots is facilitated by remodeling of their folding intermediates. *Nucleic Acids Res.* **49**, 6941–6957 (2021).
- D. Chandler, Statistical mechanics of isomerization dynamics in liquids and the transition state approximation. *J. Chem. Phys.* **68**, 2959–2970 (1978).
- P. G. Bolhuis, C. Dellago, D. Chandler, Sampling ensembles of deterministic transition pathways. *Faraday Discuss.* **100**, 421–436 (1998).
- R. Radhakrishnan, T. Schlick, Biomolecular free energy profiles by a shooting/umbrella sampling protocol, "BOLAS". *J. Chem. Phys.* **121**, 2436–2444 (2004).
- M. Karplus, J. Kuriyan, Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 6679–6685 (2005).
- L. Yang, W. A. Beard, S. H. Wilson, S. Brody, T. Schlick, Highly organized but pliant active site of DNA polymerase β : Compensatory mechanisms in mutant enzymes revealed by dynamics simulations and energy analyses. *Biophys. J.* **86**, 3392–3408 (2004).
- L. Yang *et al.*, Local deformations revealed by dynamics simulations of DNA polymerase with DNA mismatches at the primer terminus. *J. Mol. Biol.* **321**, 459–478 (2002).
- L. Yang, W. A. Beard, S. H. Wilson, S. Brody, T. Schlick, Polymerase simulations suggest that Arg258 rotation is a slow step rather than large subdomain motions *per se*. *J. Mol. Biol.* **317**, 651–671 (2002).
- Z. Lin, R. J. C. Gilbert, I. Brierley, Spacer-length dependence of programmed 1 or 2 ribosomal frameshifting on a U 6 A heptamer supports a role for messenger RNA (mRNA) tension in frameshifting. *Nucleic Acids Res.* **40**, 8674–8689 (2012).
- K. D. Mouzakis, A. L. Lang, K. A. Vander Meulen, P. D. Easterday, S. E. Butcher, HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res.* **41**, 1901–1913 (2013).
- J. Choi, S. O'Loughlin, J. F. Atkins, J. D. Puglisi, The energy landscape of 1 ribosomal frameshifting. *Sci. Adv.* **6**, eaax6969 (2020).
- X. Ou *et al.*, The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* **475**, 118–121 (2011).
- J. D. Beaudoin *et al.*, Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* **25**, 677–686 (2018).
- D. Branduardi, F. L. Gervasio, M. Parrinello, From A to B in free energy space. *J. Chem. Phys.* **126**, 054103 (2007).
- F. Pietrucci, A. M. Saitta, Formamide reaction network in gas phase and solution via a unified theoretical approach: Toward a reconciliation of different prebiotic scenarios. *Proc. Natl. Acad. Sci. U.S.A.* **112**, 15030–15035 (2015).
- 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).
- M. Zgarbová *et al.*, Refinement of the Cornell *et al.* nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902 (2011).
- D. A. Case *et al.*, Amber 2023 (University of California, San Francisco, 2023). <https://ambermd.org>. Accessed 21 April 2023.
- D. A. Case *et al.*, AmberTools. *J. Chem. Inf. Model.* **63**, 6183–6191 (2023).
- R. Salomon-Ferrer, D. A. Case, R. C. Walker, An overview of the Amber biomolecular simulation package. *WIREs Comput. Mol. Sci.* **3**, 198–210 (2013).
- D. A. Case *et al.*, The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
- J. P. Ryckaert, G. Ciccotti, H. J. C. Berendsen, Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of *n*-alkanes. *J. Comput. Phys.* **23**, 327–341 (1977).
- M. K. Scherer *et al.*, PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *J. Chem. Theory Comput.* **11**, 5525–5542 (2015).
- A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis* (John Wiley & Sons Ltd., 2001).
- G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis, F. Noé, Identification of slow molecular order parameters for Markov model construction. *J. Chem. Phys.* **139**, 015102 (2013).
- S. Yan, T. Schlick, "Heterogeneous and multiple conformational transition pathways between pseudoknots of the SARS-CoV-2 frameshift element". GitHub. <https://github.com/Schlicklab/TPS-MSM>. Deposited 27 August 2024.