Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

HHS Public Access

Author manuscript

J Mol Biol. Author manuscript; available in PMC 2026 March 10.

Published in final edited form as:

J Mol Biol. 2025 December 15; 437(24): 169512. doi:10.1016/j.jmb.2025.169512.

RNA Structural Ensemble Determinants of −1 Programmed Ribosomal Frameshifting Efficiency Across Coronavirus Evolution

Scott R. Allen[1], **Tamar Schlick**[2,3,4,5], **Alain Laederach**[1,a]

[1]Department of Biology, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA

[2]Department of Chemistry, New York University, New York, NY 10003, USA.

[3]Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA.

[4]NYU-ECNU Center for Computational Chemistry, NYU Shanghai, Shanghai 200062, P.R. China.

[5]NYU Simons Center for Computational Physical Chemistry, New York University, New York, NY 10003, USA.

Abstract

Coronavirus −1 programmed ribosomal frameshifting (−1 PRF) elements adopt complex structural ensembles that are essential for maintaining efficient frameshifting. Here we study four novel −1 PRF elements to experimentally determine and characterize their structural ensembles. Coupling this structural information with measurements of −1 PRF efficiency provides insights into the structural features responsible for refining rates of −1 PRF in better versus worse frameshifting elements (FSEs). The structure/function relationships that were previously found to govern −1 PRF efficiency in SARS-CoV-2 do not fully generalize in other FSEs. Specifically, the observation that −1 PRF efficiency increases with the number of alternative conformations is not always the case. One of the most efficient −1 PRF elements in this study, the BtRs-BetaCoV element, adopts a single conformation. We find that no single structural feature, including the relative strength of the attenuator hairpin, is predictive of frameshifting efficiency. However, we do observe a trend away from canonical stem I formation in better frameshifting elements despite reports of strong selective pressures. Finally, by swapping viral sequence elements in a modular manner, we reveal how sequence elements upstream of the attenuator hairpin substantially enhance or suppress viral frameshifting. This study provides detailed characterization of short frameshifting elements that may have biotechnological applications in packaging gene therapies or controlling stoichiometric protein expression ratios.

[a]To whom correspondence should be addressed.

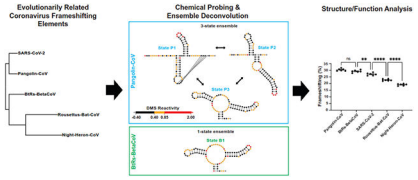
Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Graphical Abstract



Keywords

Frameshifting element; −1 programmed ribosomal frameshifting; single molecule correlated chemical probing

Introduction

Programmed ribosomal frameshifting (PRF) allows ribosomes to shift their reading frame during translation. The primary types are −1 PRF, where the ribosome moves back one nucleotide (5' direction), and +1 PRF, where it skips forward one nucleotide [1,2]. Less common +2 and −2 shifts also exist [3-5]. There are multiple mechanisms by which different forms of PRF can occur, all of which are tightly regulated [3,6-8].

In −1 PRF, regulation typically involves three RNA elements: a slippery sequence (often NNNWWWH), a spacer, and a downstream secondary structure [2,3]. Wobble position flexibility in tRNA binding at the slippery sequence encourages ribosomal slippage, while the secondary structure stalls the ribosome, increasing the chance it repositions backward by one nucleotide [2,9,10]. In contrast, +1 PRF is often triggered by ribosomal pausing due to reduced cognate tRNA availability, potentially caused by rare codons or tRNA modifications [9,11-13].

PRF enables viruses to encode multiple proteins from overlapping reading frames [14-17]. Recent work on the SARS-CoV-2 frameshift element (FSE) revealed that distinct, likely interchanging substructures are a distinct requirement of −1 PRF function. Furthermore, attempts to simplify this structural complexity through rational design approaches have frequently led to a loss of frameshifting activity [18], highlighting the delicate balance required for its mechanism.

The sensitivity observed in engineered constructs contrasts sharply with the evolutionary pressures on naturally occurring coronaviruses. For viral propagation, natural variations within the −1 PRF element must arise in ways that preserve or improve upon, rather than abolish, this essential frameshifting function. Understanding how evolution maintains this function can reveal novel insights into the requirements for −1 PRF activity. Phylogenetic analyses of coronavirus FSE structures using sequence covariation and computational modeling indicate that stem II is flexible between virus species, often involving different base pairing interactions [19]. Alternative stem I (AS1) is also a major determinant of FSE structure conformations in the Sarbecovirus subgenus when enough upstream sequence is included for its formation [17,20-23]. Functional assays of SARS-CoV-2 FSE sequences containing varying amounts of upstream sequence context reveal a functional role for

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

proximal upstream sequences, yet the structure/function relationships in these upstream regions are not fully understood [18].

Leveraging this foundation, in this work we have strategically selected for study two coronaviruses with high sequence similarity and two with lower similarity to SARS-CoV-2, ensuring representation across different viral subfamilies. Prior studies have demonstrated that coronavirus structural landscapes and frameshifting efficiency are dependent on upstream sequence context [17,19,23,24]. While much of this work has focused on long-range interactions present in intact viral genomes, proximal upstream sequences are also capable of modulating RNA structure and frameshifting efficiency [23-27]. Here we explore the structure/function relationship of proximal upstream sequence elements in five related coronaviruses. Functional assays performed on these viral elements reveal not a simple on/off state, but a spectrum or continuum of −1 PRF efficiencies. Establishing this range of functional activities provides the necessary context for investigating the specific structural differences that underpin these observed variations in frameshifting capability. These small frameshifting reporters have an added advantage that they are easily adaptable for biotechnological applications where tightly regulated protein stoichiometry or compact packaging of genetic payloads are desired.

Results

Experimental confirmation of −1 PRF activity

In this study, we investigate five distinct viral sequences (aligned in Figure 1A). To date only the SARS-CoV-2 sequence has been experimentally characterized both structurally and functionally. The other four (Pangolin-CoV, BtRs-BetaCoV, Rousettus-Bat-CoV and Night-Heron-CoV) are all predicted to be frameshifting stimulatory elements (FSEs) based on co-variation analysis and primary sequence analysis of the genomic sequences of their corresponding coronaviruses [17]. These five sequences span a broad range of evolutionary distance (78% to 99% sequence identity) within coronaviruses and will allow us to understand the functional and structural variability of FSEs and their upstream regions (Figure 1B).

In this study we consider viral sequences ~192 nucleotides in length containing the FSE, the slippery site (SS), attenuator hairpin (AH), and proximal upstream sequences. These different regions have previously been shown to regulate −1 PRF efficiency [18,26-28]. Previous studies of the SARS-CoV-2 frameshifting element identified structural ensembles containing multiple competing substructures, represented here as arc diagrams (Figure 1A) [17,18]. Previously identified stems in alternative structures of SARS-CoV-2 are labeled using RNA-As-Graphs (RAG) terminology [18]. RAG utilizes dual graphs to represent the topologies of RNA secondary structures, enabling a systematic naming of alternative conformations [29].

To experimentally confirm that each of these sequences is a functional FSE, we performed −1 PRF dual luciferase assays on each viral reporter (Figure 1C). These experiments were carried out in six biological replicates with technical triplicates. These assays are remarkably quantitative allowing us to measure subtle differences in −1 PRF efficiency. Furthermore, all

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

five wildtype sequences do result in frameshifting in our assay, with a −1 PRF efficiency ranging from 19%-30% (Figure 1C). The Night-Heron-CoV sequence is the least efficient, whereas the Pangolin-CoV and BtRs-BetaCoV have the highest frameshifting efficiency. A negative control sequence containing a stop codon preceding the slippery site exhibited less than 2% frameshifting efficiency (Figure 1C).

Comparative Structural analysis of Pangolin-CoV FSE

To begin to dissect the structure/function relationships that govern frameshifting efficiency, we performed structure probing with dimethyl-sulfate read by mutational profiling (DMS-MaP) on each viral reporter within the sequence context of the functional assay reporter plasmid. The DMS reactivity profiles strongly suggest that the viral sequences fold independently of the plasmid backbone of the dual-luciferase reporter plasmid (Supplemental Figure 1). To characterize the RNA structural ensembles we analyzed the DMS data using deconvolution and annotation of ribonucleic conformational ensembles (DANCE-MaP) [30].

With DANCE-MaP we identified three conformational states for both the Pangolin-CoV (Figure 2A) and SARS-CoV-2 (Figure 2B) RNAs. Each state corresponds to a unique pattern of DMS reactivity for the Pangolin-CoV and SARS-CoV-2 SS and FSE regions (Figure 2C,D). Hereafter, the 87 nucleotide region comprising of the FSE, the SS, and the codon immediately preceding the SS will be referred to as the 87nt-FSE region. When we compute the pairwise correlation coefficients between these different reactivity patterns for the 87nt-FSE region (Figure 2E), we find that each state identified in the SARS-CoV-2 ensemble is highly correlated with a corresponding state in the Pangolin-CoV ensemble. For example, Pangolin-CoV state P1 is highly correlated with SARS-CoV-2 state S3 in this region, with a correlation coefficient of 0.94 (Figure 2E). Furthermore, the DMS reactivities for the correlated Pangolin-CoV and SARS-CoV-2 states are nearly superimposable as illustrated by the overlapping skyline profiles (Figure 2F).

When modelling the secondary structures using DMS reactivities as pseudo free energy constraints we observe a high degree of variability in the forward predicted models between biological replicates, even when DMS reactivities are within error. We exemplify this for SARS-CoV-2 state 2 where the biological replicate DMS reactivities are highly correlated ($r >= 0.97$; Supplemental Figure 2A-C). For modelling the well characterized SARS-CoV-2 87nt-FSE we find that fitting the DMS reactivities to existing models of the SARS-CoV-2 87nt-FSE region with receiver operating characteristic (ROC) curves is more robust across biological replicates (Supplemental Figure 2D,E). The best-fit structure models based on area under the ROC curve (AUROC) are highly consistent with the measured DMS reactivities (Supplemental Figure 2F-H) and are also similar to forward predicted models in at least one biological replicate (compare top panel in Supplemental Figure 2C with Supplemental Figure 2F). We expanded this AUROC analysis to each DANCE-deconvoluted state of SARS-CoV-2 and also Pangolin-CoV based on their aforementioned near identical sequences and DMS reactivities (Supplemental Figure 3).

Previous computational and DMS-MaP analysis of the 87nt-FSE region of SARS-CoV-2 identified five major alternative conformations for this region, which are classified as

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

3_6, 3_5, 3_3, 2_1 and 3_5* (Supplemental Figure 3A) using the RNA as Graphs (RAG) approach [18]. Given each of these known conformations we can compute an ROC curve for each reactivity profile to each corresponding structure [31]. A higher area under the ROC curve (AUROC) indicates higher agreement between the DMS reactivity profile and a given structure (Supplemental Figure 3B,C). We can thus use the AUROC values to assign each SARS-CoV-2 and Pangolin-CoV conformational state to the appropriate previously described conformation (summarized in Supplemental Figure 3D) without relying on forward predicted thermodynamic modelling. These data are consistent with previous findings for SARS-CoV-2 and suggest that for these two viral sequences, the conformational ensemble is conserved. These data also show that DANCE-MaP deconvolution is quantitative and identifies conserved structural signatures in closely related FSEs.

Given that several groups have reported long range structure interactions involving the SARS-CoV-2 FSE [23-25], it is important to note that we use only a 192 nucleotide fragment of the SARS-CoV-2 genome in our reporter. It has been shown that minimal FSE reporters can vary dramatically in their structure and DMS reactivity profiles compared to those with more sequence context or in cellular infection models of SARS-CoV-2 [23]. To determine whether the structure adopted by the reporter sequence used here has biological relevance to intact SARS-CoV-2 we compared the population averaged DMS reactivities of the reporter used in this study to those reported by Lan et. al., 2022 [23]. We observe a high correlation between the DMS reactivities of the reporter used here and SARS-CoV-2 infection models in Vero cells ($r = 0.84$; Supplemental Figure 4A) and Huh7 cells ($r = 0.83$; Supplemental Figure 4B), indicating that our in vitro reporters are representative of the structural reality in cells. Furthermore, we note similarly high correlation when comparing the DMS reactivities measured in this study with those of longer *in vitro* reporters with the strongest correlation being to *in vitro* transcribed whole SARS-CoV-2 genome ($r = 0.87$; Supplemental Figure 4C-D). We find no correlation between the DMS reactivities of the reporter used in this study and a minimal 88 nucleotide reporter shown to adopt a different structure from that found in intact virus ($r = -0.01$; Supplemental Figure 4F), suggesting our choice of reporter lengths captures more essential structural features of the region than other minimal reporters.

Ensemble analysis of BtRs-BetaCoV, Rousettus-Bat-CoV and Night-Heron-CoV viral sequences

Having established with dual luciferase assays that all five viral sequences exhibit frameshifting activity (albeit with different levels of efficiency, Figure 1C) we now aim to compare their structural ensembles. We performed DANCE-MaP deconvolution on DMS data we collected for these three additional viral sequences. Interestingly none of these sequences deconvolved into more than two states, with the BtRs-BetaCoV having only one state (Figures 3A-3C). This was surprising, as BtRs-BetaCoV also has one of the highest frameshifting efficiencies (29%, Figure 1C). Previous work on the SARS-CoV-2 viral sequence has found that frameshifting efficiency generally increases with the number of states [18]. However, these data suggest that this property may not be generalizable to other coronaviruses.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Focusing in on the 87nt-FSE region, we observe that BtRs-BetaCoV retains high sequence similarity with SARS-CoV-2 and Pangolin-CoV (differing by <5 nucleotides, Figure 1A) and displays near identical DMS reactivities to Pangolin-CoV state 3 (Figure 3D). This suggests that it likely adopts the same 3_5* structural conformation as Pangolin-CoV state 3 (Supplemental Figure 5). Rousettus-Bat-CoV and Night-Heron-CoV, however, differ substantially in primary sequence, both from the other coronaviruses in this study and from each other (Figure 1A). Rousettus-Bat-CoV adopts two structural conformations (Figure 3B) which contain distinct conformational states in the 87nt-FSE region (Figure 3E). Night-Heron-CoV also adopts two structural conformations (Figure 3C), however, these states adopt identical conformations for the 87nt-FSE region (r = 0.95 ; Figure 3F).

Rousettus-Bat-CoV and Night-Heron-CoV differ too much in primary sequence from the other three coronaviruses in this study to assign to existing structure models using AUROC (Figure 1A). Thus, we use DMS-informed thermodynamic modeling with ShapeKnots [32] to generate structure models for these viral elements (Figure 3H-J) while acknowledging the caveats of forward predictions. Rousettus-Bat-CoV state R1 adopts a simple 3-stem conformation 3_1 (Figure 3H), and state R2 adopts a 3_5 conformation (Figure 3I) previously observed in SARS-CoV-2 (Supplemental Figure 3A). The Night-Heron-CoV 87nt-FSE region also adopts a simple 3_1 conformation (Figure 3J). Notably, we observe covarying base pairs for both stem I and stem III in the Rousettus-Bat-CoV structure models, and for both stem I and 3_1 stem II in the Night-Heron-CoV model (Supplemental Table 1). This suggests these structures may have functional relevance and be under purifying selection. We also observe covarying base pairs for SARS-CoV-2 stem I, AS1, and 3_6 stem II (Supplemental Table 1) in line with previous studies [17,25].

Overall, the five coronaviruses in this study appear to adopt a limited number of conformations and converge on a limited set of substructures (Table 1). Interestingly, the worse frameshifting Rousettus-Bat-CoV and Night-Heron-CoV still adopt many of the same substructures as their better frameshifting counterparts, and even some of the same conformations (Figure 3I). All structural conformations across the coronaviruses in this study contain either stem I or stem I*, and all except Night-Heron-CoV contain a clear stem III (Table 1). Additionally, all conformations have some variation of stem II (Table 1), all of which involve the same group of nucleotides on the 3’ end of the stem (Figure 2G, Figure 3G-J). These nucleotides are highly conserved across all coronaviruses in this study (Figure 1A).

Correlated chemical probing analyses provide modeling-agnostic insights into structural determinants of −1 PRF efficiency

Deconvolution of correlated mutations using DANCE-MaP identified conserved alternative conformations in the 87nt-FSE region of the coronavirus sequences studied here. However, both forward predicted thermodynamic models and using AUROC to fit DMS reactivities to pre-existing models of minimal reporter elements have limitations. AUROC depends on previously characterized structures that may be inaccurate or not fully represent the deconvolved ensemble. Both thermodynamic modeling (Supplemental Figure 2C) and AUROC (Supplemental Figure 3D) can be sensitive to small changes in the data, although

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

AUROC was more robust in our analyses (Supplemental Figure 2D, E). An alternative analysis of the data using pairing ascertained from interacting RNA strands measured by mutational profiling (PAIR-MaP) can reveal direct pairing interactions, or “PAIRs” with high degrees of confidence [33]. PAIRs are computed based on correlated chemical probing data and can be classified based on their statistical significance. Primary PAIRs have the highest confidence of representing base-pairing events, followed by secondary PAIRs and complimentary PAIRs. Here we focus on the higher confidence classes of primary and secondary PAIRs. We computed PAIRs for the 87nt-FSE regions of Night-Heron-CoV and BtRs-BetaCoV given that both adopt a single conformation in this region yet differ dramatically in −1 PRF efficiency. The computed PAIRs suggest that although the Night-Heron-CoV 87nt-FSE region may fold independently (Figure 4A) there are multiple PAIRs indicating interactions between the 87nt-FSE region and upstream sequence elements in BtRs-BetaCoV (Figure 4B).

We next analyzed the PAIRs across the upstream and 87nt-FSE regions. The better frameshifting elements all contain either primary or secondary PAIRs supporting AS1 (Figure 4E-G), whereas neither of the worse frameshifting elements do (Figure 4C,D). This is consistent with phylogenetic studies suggesting that AS1 is conserved among the Sarbecovirus subgenus but not in other coronaviruses [17]. It further suggests a functional role for AS1 or perhaps upstream-FSE interactions in general for determining −1 PRF efficiency. The two best frameshifting elements, Pangolin-CoV and BtRs-BetaCoV, have PAIRs supportive of additional upstream-FSE interactions (Figure 4E,F), further suggesting that upstream-FSE interactions may bolster −1 PRF efficiency. Notably this trend holds true even among the worse frameshifting elements. Night-Heron-CoV which induces the least −1 PRF contains no PAIRs supporting upstream-FSE interactions (Figure 4D), while the slightly better frameshifting Rousettus-Bat-CoV contains a single secondary PAIR offering weak support for an upstream-FSE interaction (Figure 4C).

Other groups have also observed AS1 in structural ensembles of SARS-CoV-2 in cellular infection models [23,25]. We used AUROC analysis to see if the DMS reactivities for any of the SARS-CoV-2 and Pangolin-CoV states are consistent with those AS1 containing models, and whether those models are supported by PAIRs interactions (Supplemental Figure 6). We find that SARS-CoV-2 state S1 is best fit to conformation A reported by Morandi et. al. and state S2 is best fit to conformation B reported by Morandi et. al. (Supplemental Figure 6A) [25]. SARS-CoV-2 state S3 is still best fit to the cryo-EM supported conformation 3_6 [15] with much better AUROC values than the cellular infection models (Supplemental Figure 6A). It should be noted that the cluster 1 conformation reported by Lan et. al. contains additional 3’ sequence and the cluster 2 conformation contains long-range interactions that are not possible to observe in our reporter. The best fit models by AUROC analysis are highly consistent with the observed DMS reactivities for SARS-CoV-2 and several of the stems are supported by PAIRs (Supplemental Figure 6B-D). The DANCE-deconvoluted DMS reactivities reported here also correlate well with the DRACO-deconvoluted DMS reactivities reported by Morandi et. al. for conformation A ($r = 0.63$; Supplemental Figure 6E) and conformation B ($r = 0.82$; Supplemental Figure 6F). The DMS reactivities for equivalent SARS-CoV-2 and Pangolin-CoV states are also highly correlated across the expanded sequence context covered by these structure models ($r >= 0.94$; Supplemental

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Figure 6G). Taken together these analyses suggest that conformations A and B reported by Morandi et. al. are present in the structural ensembles of both the SARS-CoV-2 and Pangolin-CoV reporters used in this study.

Examining PAIRs supportive of structural elements fully contained in the 87nt-FSE region also reveals differences between better and worse frameshifting elements. Worse frameshifting elements contain strong primary PAIRs in support of stem I (Figure 4C, D), whereas none of the better frameshifting elements contain any stem I-supportive PAIRs. Instead, the better frameshifting elements contain PAIRs supportive of the competing stem I* in two out of three viral sequences (Figure 4E, G). All the viral sequences contain multiple secondary PAIRs in support of stem III, except for the worst frameshifting Night-Heron-CoV, which contains a single secondary PAIR interaction in support of either a broader stem III or a previously uncharacterized variant of stem II (Figure 4D).

A known major regulator of frameshifting efficiency in coronaviruses is the attenuator hairpin (AH). Pangolin-CoV and BtRs-BetaCoV are tied for having the highest frameshifting efficiency out of the reporters tested (Figure 1C). As one might expect, Pangolin-CoV lacks PAIRs supporting the negative-regulating AH (Figure 4E). However, BtRs-BetaCoV has strong primary PAIRs supporting an extended (stronger) AH compared to that of SARS-CoV-2 (see extension past AH region in Figure 4F). This is surprising because BtRs-BetaCoV has higher frameshifting efficiency than SARS-CoV-2 (Figure 1C). Additionally, one of the worse frameshifting elements, Rousettus-Bat-CoV, lacks PAIRs supportive of the attenuator hairpin (Figure 4C) and instead has multiple secondary PAIRs interactions supporting the upstream hairpin which straddles the slippery site (Figure 4C). We anticipated that a worse frameshifting element might have an extended AH, and a better frameshifting element might have a weakened AH, yet these data suggest a less clear relationship between the AH formation and −1 PRF efficiency among coronaviruses outside of the well-studied SARS-CoV-2.

There are numerous PAIRs supporting self-contained structures in the upstream sequences of every viral sequence (Figure 4C-G). The upstream sequence elements of SARS-CoV-2 and Night-Heron-CoV appear to fold the most independently given that they lack PAIRs supporting upstream-FSE interactions upstream of AS1 (Figure 4D, G). To explore the relationship between upstream sequence elements and −1 PRF efficiency, we swapped the reporter sequences upstream of the AH between different coronavirus sequences.

Proximal sequence elements contain novel regulators of −1 PRF efficiency

We generated mutants where we swapped the sequences upstream of the well-defined SARS-CoV-2 attenuator hairpin between the better frameshifting SARS-CoV-2 and the worse frameshifting Night-Heron-CoV (Figure 5A). Dual luciferase assays comparing the frameshifting efficiency of these mutants to their wildtype counterparts reveal a strikingly clean relationship. Replacing the wildtype upstream sequence of the worse frameshifting Night-Heron-CoV with that of the better frameshifting SARS-CoV-2 results in dramatically enhanced frameshifting efficiency (Figure 5B). Conversely, replacing the wildtype upstream sequence of the better frameshifting SARS-CoV-2 with that of the poorer frameshifting Night-Heron-CoV leads to dramatically reduced frameshifting efficiency (Figure 5B).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

The difference in frameshifting efficiency achieved by swapping the upstream sequence elements accounts for most of the difference between wildtype SARS-CoV-2 and wildtype Night-Heron-CoV (Figure 5B). We next compared the PAIRs interactions between the SARS-CoV-2 and Night-Heron-CoV swap mutants and their respective wildtype reporters (summarized in Table 2).

We were surprised to find no differences in PAIRs interactions between the improved Swap-SN mutant and wildtype Night-Heron-CoV reporters across the retained Night-Heron-CoV sequence region containing the AH and 87nt-FSE (Figure 5C). This is despite the increase in −1 PRF efficiency. Analysis of the DMS reactivities in the retained Night-Heron-CoV sequence region revealed the improved Swap-SN mutant gains a new structural conformation in the 87nt-FSE region (Supplemental Figure 7A-C). The addition of conformational states correlating with the increased frameshifting efficiency of the Swap-SN mutant is consistent with previously observed trends in SARS-CoV-2 [18].

The new 87nt-FSE conformation in the improved Swap-SN mutant likely has a weakened or abolished stem I as revealed by enhancements in this region using deltaDMS analysis (Supplemental Figure 7D, E). This would be consistent with our previous observations that better frameshifting elements tend to disfavor stem I formation in favor of the competing AS1 or stem I* as supported by both structure models (Table 1) and PAIRs interactions (Figure 4). Notably, we were able to increase −1 PRF yet observed no AS1- or stem I*-supportive PAIRs (Figure 5C), further indicating the presence of a novel −1 PRF regulatory element in the upstream sequence region.

When comparing PAIRs interactions between the worsened Swap-NS mutant and the wildtype SARS-CoV-2 reporters, we observe several changes (Figure 5D). The most dramatic of these differences is the apparent extension/strengthening of the attenuator hairpin (AHext in Figure 5D) which would be consistent with the negative-regulatory role of the AH and observed reduction in −1 PRF efficiency in the mutant. Although AS1 has been disrupted in the Swap-NS mutant, there are numerous secondary PAIRs supporting an AS1-like stem with the same 3’ nucleotides pairing with a set of nucleotides further upstream (Figure 5D). Whereas the loss of AS1 would be consistent with observed patterns between better and worse frameshifting elements (Figure 4C-G), this AS1-like interaction makes the structure/function interpretation in the case of the Swap-NS mutant unclear. The worsened Swap-NS mutant also loses a stem I* secondary PAIR and gains a stem I secondary PAIR compared to wildtype (Figure 5D). This follows the observed patterns regarding stem I formation in better versus worse frameshifting elements (Table 1, Figure 4) in agreement with the reduced frameshifting efficiency. The PAIRs interactions are consistent with comparisons of the DMS reactivities in the retained SARS-CoV-2 sequence region between the wildtype and swap mutant reporter. The DMS reactivity profiles indicate a loss of the DMS reactivity pattern associated with the 2_1 conformation (which contains stem I*) in favor of a higher proportion of DMS reactivities associated with the 3_6 conformation (which contains stem I) in the 87nt-FSE structure ensemble (Supplemental Figure 8).

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Taken together, the analyses of the swap mutants reveal the presence of potent −1 PRF regulatory elements in the sequence region just proximal of the AH region. These elements are contained within the 80 nucleotides upstream of the AH region and may exert most of their effect through direct base pairing interactions (Figure 5D). Importantly, these regulatory effects are independent of AS1 or any previously characterized structural elements (Figure 5C).

Discussion

Thermodynamic structure models informed by chemical probing data offer efficient, cost-effective secondary structure solutions in many instances where characterization of novel RNA elements is desired. When dealing with biologically functional structured RNA elements that are both highly conserved at the level of primary sequence and adopt complex structural ensembles, however, thermodynamic modeling programs can be overly sensitive to minor changes in reactivities. This sensitivity over exaggerates differences between structure models in ways inconsistent with comparisons of the underlying raw reactivities (Supplemental Figure 2A-C). Mutagenesis studies to correct or validate such models can quickly become cost and resource prohibitive to achieve as scale increases, and are inherently challenging when dealing with ensembles of structures. Approaching the study of such structured elements first with a series of thermodynamic modeling-agnostic approaches can efficiently and effectively reveal which structured elements are likely to be conserved or to differ between related RNA species. Importantly, in a system as thoroughly studied at the SARS-CoV-2 FSE the approach we used here allows us to understand our data in the context of the existing literature.

When modeling specific structured elements in the context of large RNAs it is sometimes appropriate or necessary to model discreet regions of RNA sequence independently. This study highlights the usefulness of single-molecule correlated chemical probing analyses such as PAIR-MaP in determining which sequence regions to model independently. In the case of coronavirus −1 PRF elements, previous studies have independently modeled the 87nt-FSE region of larger RNA reporters [18]. PAIR-MaP analyses presented here partially support this approach but show that conserved structure interactions are ignored when independently modeling this region. Namely, stem I* modeled in the 2_1 and 3_5* conformations observed in longer SARS-CoV-2 sequences [18] likely represent a mixed population of stem I* and AS1. PAIRs supporting both stem I* and AS1 are observed in the better frameshifting elements (Figure 4E-G) with favor leaning towards AS1 formation.

Prior works have suggested clear functional roles for the structural plasticity [17-20,34], the presence and strength of the attenuator hairpin [27], and the presence of AS1 in modulating −1 PRF efficiency [28]. The current study does not contradict previous findings but reveals limitations in using any individual structural feature to predict relative frameshifting efficiency. In the viruses studied here, the number of conformational states in the 87nt-FSE sometimes correlates with −1 PRF but alone is not a good predictor of −1 PRF efficiency as clearly demonstrated by BtRs-BetaCoV (Table 1). Similarly, the presence, absence, and predicted strength of the attenuator hairpin is not a good predictor of −1 PRF (Figure 4C, F). The formation of AS1 may increase frameshifting efficiency or may have only appeared

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

to do so in our system due to the high level of sequence conservation between better frameshifters (Figure 1A, Figure 4E-G) and conservation of AS1 within the Sarbecovirus subgenus [17].

Focusing on individual stems within the frameshift stimulation element (FSE) itself also reveals structure/function correlations. Viruses with higher rates of −1 PRF tend to disfavor formation of stem I in favor of stem I* or AS1 in this study (Table 1 and Figure 4). This trend holds true for the swap mutants as well. The only new 87nt-FSE state observed in the improved Swap-SN mutant disfavors stem I formation (Supplemental Figure 7E), and the worsened Swap-NS mutant favors stem I formation (Figure 5D). This is in contradiction with prior covariation and phylogenetic analyses suggesting strong selection for stem I formation [17,21]. Stem I is also observed in the pseudoknotted cryo-EM structure of the SARS-CoV-2 frameshifting element [15]. Notably, stem I* and AS1 both involve nucleotides upstream of those involved in forming stem I and would necessarily be unwound before the ribosome encounters the stem I forming region of the FSE. Together these findings suggest that structural rearrangements to actively form stem I in the presence of proximal upstream translating ribosomes may be a mechanistic feature of more efficient coronavirus frameshifting, whereas having stem I preformed in the resting state results in less frameshifting. In fact, such dynamic structural rearrangements have frequently been proposed for transitioning from AS1 to stem I containing conformations in the presence of translating ribosomes [20,28]. However, more extensive studies are needed to refine the mechanism by which dynamic structural rearrangements lead to incredibly reproducible −1 PRF rates. It has been postulated that rapid structural rearrangements lead to fixed fractions of states conducive (pseudoknotted) or not conducive (not pseudoknotted) to frameshifting, resulting in the observed rates of −1 PRF [28]. Yet experimental evidence from independent groups repeatedly demonstrates that structural conformation stabilizing mutants abolish frameshifting efficiency, even when the presumably active pseudoknotted conformation is stabilized [18,28].

In contrast to stem I, the differences in stem II formation between better and worse frameshifters offer less clear structure/function interpretations. All coronaviruses in this study appear to have some variant of stem II with the same 3’ nucleotide anchor. The nucleotides involved in this 3’ anchor are also well conserved among the viruses studied (Figure 1A). The structure models suggest that stem II more commonly forms multihelix junctions or bulged stems in the better frameshifting elements (Figures 2G and 3G), and short simple stems in worse ones (Figure 3H, J). These observations are in agreement with phylogenetic analysis and computational modeling studies that indicate flexible stem II formation across coronavirus evolution and a trend towards simpler multi-stem FSE conformations as reporter sequence length increases [19,20]. Indeed, Rousettus-Bat-CoV was previously reported to contain mutations favoring 3_5 stem II formation and we observe 3_5 stem II in our DMS informed model for state R2 (Figure 3I) [19]. However, the lack of PAIRs supporting stem II formation in any of the coronaviruses except Night-Heron-CoV suggests the structure model predictions for stem II should be interpreted with caution.

All the coronaviruses in this study also form some variant of stem III or, in the case of Night-Heron-CoV, a new variant of stem II that resembles stem III (Figure 4D). This is

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

consistent with prior covariation analysis of coronaviruses suggesting that stem III is less conserved in the FSE, often containing deletions that might result in different lengths or locations of stem III [17]. Indeed, we observe variations in stem III length and location in both the structure models (Figures 2G and 3G-J) and PAIRs interactions (Figure 4). There is a trend towards weaker stem III formation in worse frameshifting elements.

Primary sequence features can also impact frameshifting efficiency. One of the best examples of this is the sequence of the E-site codon immediately preceding the slippery site [35,36]. The E-site codon does differ between the better and worse frameshifting elements in this study with the better frameshifting elements all having GUU and the worse ones having either CUU (Rousettus-Bat-CoV) or UUA (Night-Heron-CoV). Although E-site codon identity may contribute to some of the differences observed in better versus worse frameshifting elements in this study it cannot fully explain the observed differences. The swap mutants achieve large differences in frameshifting efficiency without altering the identity of the E-site codon.

The functional assays with the swap mutants demonstrate the importance of proximal upstream sequence elements on frameshifting efficiency. Much attention has been given to the attenuator hairpin, alternative stem I (AS1), and long-range RNA interactions in coronavirus frameshifting [24,37]. Long-range RNA interactions have also been a focus of studies in other viral frameshifting elements [38]. This study highlights the importance of considering proximal upstream sequence elements beyond the attenuator hairpin and AS1. Further exploring the structural and functional consequences of this sequence region will be important in future studies.

Our work demonstrates that nature has fine-tuned −1 PRF efficiency in different viral sequences. Interestingly no single structural or sequence feature alone can predict −1 PRF efficiency, but rather their combined effects lead to remarkably reproducible stoichiometric expression ratios of upstream and downstream ORFs. This work not only sheds light on the fundamental mechanisms of −1 PRF but also provides a general framework and toolset with which to fine tune the stoichiometric expression ratios of any two encoded proteins using the minimal frameshift elements characterized here.

Materials and Methods

Sequences and reporter design

Coronavirus sequences were obtained from GenBank for SARS-CoV-2 (Wuhan-Hu-1, NC_045512.2), Pangolin-CoV (MP789, MT121216.1), BtRs-BetaCoV (BtRs-HuB2013, KJ473814.1), Rousettus-Bat-CoV (GCCDC1 356, NC_030886.1), and Night-Heron-CoV (HKU19-6918, NC_016994.1). Sequence alignments for the viral insert sequences for −1 PRF reporter design were obtained from Schlick et. al., 2021, and were verified against the genomic sequences for native reading frame [17]. Swap mutants were designed by swapping sequences upstream of the attenuator hairpin-aligned region and nucleotides were added where necessary to maintain reading frame. Reporter sequences were synthesized and cloned by GenScript into the pJD2257 plasmid backbone using BamHI and SalI cloning sites. The pJD2257 plasmid is a 0-frame readthrough dual-luciferase control plasmid,

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

obtained as a kind gift from Professor Jonathan D. Dinman, University of Maryland. Purified plasmids were sequence verified by whole plasmid sequencing (Eurofins, USA). Viral insert sequences for all wildtype and mutant reporters are provided in Table 3.

Phylogenetic Tree

Phylogenetic trees were generated using the full genomic sequences of each virus and the Clustal Omega multiple sequence alignment web portal with default parameters (https://www.ebi.ac.uk/jdispatcher/msa/clustalo?stype=rna).

Dual-luciferase assays of −1 PRF

Dual-luciferase assays were performed as in Dey et. al., 2024 [18]. Briefly, Hek293 cells were transfected with a 0-frame readthrough control pJD2257 plasmid or coronavirus FSE dual-luciferase reporter plasmid using TransIT®-Oligo Transfection Reagent (Mirus Bio). At 24 hours post transfection Hek293 cells were washed with 1x PBS and lysed with 1x passive lysis buffer (Promega). Lysates were briefly centrifuged to remove cellular debris and were assayed in a 96 well-plate. Luciferase activity was measured using the Promega Dual-Luciferase Reporter Assay System (Promega, cat. E1960) and a CLARIOstar plate reader (BMG LABTECH). The percentage of −1 PRF was calculated as the ratio of firefly to *Renilla* luciferase for each reporter, normalized to that of the control pJD2257 reporter (which was set to 100%). All reporters in this study (except Neg. Ctrl, which was added later) were analyzed in each biological replicate, with technical triplicates, allowing for direct comparisons between all reporters. All reporters had 6 biological replicates, each with technical triplicates, measured across at least 3 separate days. Each biological replicate contained technical triplicates of the control pJD2257 plasmid for normalization. Biological replicates in this study indicate different cell passages and technical replicates indicate independently transfected wells on a 96 well plate. Biological replicate −1 PRF values were averaged from the technical replicates. ANOVA ($F=873$, $p<0.0001$) was followed by Tukey's multiple comparisons test to compute all pairwise comparisons between all reporters in this study. Selected results from this analysis are shown in Figure 1C and Figure 5B.

In vitro chemical probing of coronavirus reporter sequences

RNA used for chemical probing was prepared as follows. Reporter plasmids used in dual-luciferase assays were digested with XbaI, and linearized plasmid was purified using a Monarch® DNA Cleanup Kit (NEB, USA). Linearized plasmid was used as a template for *in vitro* transcription using HiScribe® T7 High Yield RNA Synthesis Kit (NEB, USA) for 16 h at 37°C. The RNA was treated with Turbo DNase (Invitrogen, USA) and purified using Monarch® Spin RNA Cleanup kit (NEB, USA). Transcript size and purity was verified by capillary electrophoresis on a TapeStation (Agilent, USA).

DMS probing reactions were carried out in bicene-buffered conditions previously optimized for DMS reactivity against all four nucleotides [39]. RNA was diluted to a total of 5 ug in 25 uL H_2O, heat denatured at 98°C for 1 minute, and snap cooled on ice for 1-2 minutes. 25 uL of 2x folding buffer (600 mM Bicene pH 8.3, 200 mM NaCl, 10 mM $MgCl_2$) were added and the RNA was folded at 37°C for 30 minutes. 9 uL of folded RNA were combined with 1

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

uL of DMS solution (1.7 M in NS) or 1 uL of vehicle (NS) as in Ehrhardt and Weeks 2020 [40]. NS is comprised of 1:2 (v/v) nitromethane/sulfolane. Reactions were incubated at 37°C for 6 minutes, quenched with equal volumes of 25% 2-mercaptoethanol, and kept on ice. Modified and unmodified RNA was purified using Cytiva microspin G-50 columns (Fisher Scientific, USA). Two independent replicates of *in vitro* transcription and DMS probing reactions were carried out for each reporter.

Library construction, sequencing, and data processing

Mutational profiling reverse transcription (MaP-RT) was performed as in Smola et. al., 2015 [41] using random nonamers (NEB, Random Primer 9) and SuperScript II RT enzyme (Thermo Fisher Scientific) under error-prone conditions. Briefly, 8.8 uL of purified RNA were combined with 1 uL of 200 ng/uL random nonamer and 2 uL of 10 mM dNTPs. This mixture was heat denatured at 98°C for 1 minute, and snap cooled on ice for 1-2 minutes. 2 uL of freshly prepared 10x NTP-minus first strand buffer (0.5 M Tris pH 8.0, 0.75 M KCl, 0.1 M DTT), 4 uL of 5 M betaine, and 1.2 uL of 100 mM $MnCl_2$ were added and the reaction was incubated at 25°C for 2 minutes. 1 uL of either SuperScript II RT enzyme or H2O (for NRT controls) was added and reactions were incubated in a thermocycler for 10 minutes at 25°C, 90 minutes at 42°C, followed by ten cycles of 50°C for 2 minutes then 42°C for 2 minutes, then incubated at 70°C for 10 minutes to deactivate the RT enzyme. Reactions were cleaned with Cytiva microspin G-50 columns.

For constructing next-generation sequencing libraries, purified cDNA was PCR amplified using primers directed against the pJD2257 plasmid backbone (primer sequences in Table 4) and NEB Q5® Hot Start High-Fidelity DNA Polymerase (NEB). Unique barcodes were introduced with secondary PCR using NEBNext® Multiplex Oligos for Illumina® (NEB). Libraries were purified using AMPure-XP magnetic beads, quality checked on an Agilent TapeStation, and quantified using a Qubit 1x dsDNA HS kit (Invitrogen). Libraries were sequenced as 2 x 300 paired end reads on a NextSeq1000.

Sequencing reads were processed with ShapeMapper2 [42] (version 2.2.0) to calculate mutation frequency in both modified (DMS) and unmodified (NS) samples using the following flags: --dms, --amplicon, --primers, --output-parsed-mutations. Generation of all profile, skyline and arc plots, and linear regression and ROC analyses were performed in RNAvigate (version 1.0.0) [31].

DANCE-MaP analysis on coronavirus reporter sequences

DANCE-MaP [30] is a single-molecule chemical probing strategy that extracts large amounts of information from single chemical probing experiments. This includes DMS reactivity profiles for individual structure states in an RNA ensemble as determined using a maximum likelihood framework and modified Bernoulli mixture model. Parsed mutation files limited to the coronavirus insert sequences (for both modified and unmodified samples) from ShapeMapper2 [42] were input to the DANCE-MaP pipeline to determine the number of states in each coronavirus structure ensemble and to generate DMS reactivity profiles for each state.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

ShapeKnots

Normalized DMS reactivities generated by the DANCE-MaP pipeline for Rousettus-Bat-CoV and Night-Heron-CoV states were trimmed to include only the 87nt-FSE region. These reactivities were used as pseudo free energy constraints in ShapeKnots [32] using the --DMSNT flag to predict minimum free energy models for each state.

PAIR-MaP analysis on coronavirus reporter sequences

PAIR-MaP uses single-molecule correlated chemical probing data to detect direct base-pairing interactions in RNA molecules with high resolution and confidence [33]. Population averaged DMS reactivities and parsed mutation files limited to the coronavirus insert sequences (for both modified and unmodified samples) from ShapeMapper2 [42] were input to PairMapper (version 1.2). Resulting primary and secondary PAIRs were visualized in arc plots using RNAvigate [31].

DeltaDMS analysis on the Swap-SN mutant

DeltaDMS analysis is the DMS probing equivalent of deltaSHAPE analysis [43], and identifies nucleotides with significantly different DMS reactivities between two conditions. DeltaDMS was computed by first averaging the DMS reactivity profiles between replicates for wildtype Night-Heron-CoV (population averaged reactivities) and Swap-SN mutant state SN3. Per nucleotide error rates were calculated as the standard deviation between replicates. Importantly, these files were trimmed to only contain the retained Night-Heron-CoV nucleotides. These reactivities were imported into RNAvigate [31] (version 1.0.0) and analyzed using the built-in DeltaSHAPE analysis pipeline.

Covariation analysis

We built covariance models (CMs) for previously reported 87nt-FSE SARS-CoV-2 conformations 2_1, 3_3, 3_5, 3_5*, and 3_6 as well as previously reported stems AS1, AH, and UH using the cmbuild module of Infernal (v1.1.4) [44] using 1248 unique sequences for RNA homologs in a list of 1248 non-redundant coronavirus genomes manually curated from the GISAID and ViPR databases by Schlick et. al. [17]. Conformations 3_6 and 3_3 were split into two CMs with a separate structure model for the pseudoknot stem due to the inability of Infernal covariance models to handle pseudoknots. We similarly built CMs for 87nt-FSE Rousettus-Bat-CoV conformations 3_1 and 3_5 and Night-Heron-CoV conformation 3_1 (Figure 3H-J). The resulting CMs were calibrated using the cmcalibrate module of Infernal. The multiple sequence alignment files were input to the R-Scape web server with the “evaluate a given structure” option and default parameters (E threshold = 0.05) to identify covarying base pairs (Supplemental Table 1) [45-47].

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by U.S. National Institutes of Health grants NHLBI R01 HL111527 and NIGMS R35 GM140844 to A.L. and National Science Foundation Award DMS −2151777 from the Division of Mathematical

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Sciences to T.S. and A.L, National Institutes of Health R35GM122562 from the National Institute of General Medical Sciences to T.S., and National Science Foundation Award CHE-2330628 to T.S. and A.L.

References:

[1]. Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV, Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use, Nucleic Acids Res. 44 (2016) 7007–7078. 10.1093/nar/gkw530. [PubMed: 27436286]

[2]. Dinman JD, Mechanisms and implications of programmed translational frameshifting, WIREs RNA 3 (2012) 661–673. 10.1002/wrna.1126. [PubMed: 22715123]

[3]. Riegger RJ, Caliskan N, Thinking Outside the Frame: Impacting Genomes Capacity by Programmed Ribosomal Frameshifting, Front. Mol. Biosci 9 (2022) 842261. 10.3389/fmolb.2022.842261. [PubMed: 35281266]

[4]. Fang Y, Treffers EE, Li Y, Tas A, Sun Z, Van Der Meer Y, De Ru AH, Van Veelen PA, Atkins JF, Snijder EJ, Firth AE, Efficient −2 frameshifting by mammalian ribosomes to synthesize an additional arterivirus protein, Proc. Natl. Acad. Sci 109 (2012). 10.1073/pnas.1211145109.

[5]. Li Y, Firth AE, Brierley I, Cai Y, Napthine S, Wang T, Yan X, Kuhn JH, Fang Y, Programmed −2/−1 Ribosomal Frameshifting in Simarteriviruses: an Evolutionarily Conserved Mechanism, J. Virol 93 (2019) e00370–19. 10.1128/JVI.00370-19. [PubMed: 31167906]

[6]. Napthine S, Ling R, Finch LK, Jones JD, Bell S, Brierley I, Firth AE, Protein-directed ribosomal frameshifting temporally regulates gene expression, Nat. Commun 8 (2017) 15582. 10.1038/ncomms15582. [PubMed: 28593994]

[7]. Hill CH, Cook GM, Napthine S, Kibe A, Brown K, Caliskan N, Firth AE, Graham SC, Brierley I, Investigating molecular mechanisms of 2A-stimulated ribosomal pausing and frameshifting in *Theilovirus*, Nucleic Acids Res. 49 (2021) 11938–11958. 10.1093/nar/gkab969. [PubMed: 34751406]

[8]. Li Y, Treffers EE, Napthine S, Tas A, Zhu L, Sun Z, Bell S, Mark BL, Van Veelen PA, Van Hemert MJ, Firth AE, Brierley I, Snijder EJ, Fang Y, Transactivation of programmed ribosomal frameshifting by a viral protein, Proc. Natl. Acad. Sci 111 (2014). 10.1073/pnas.1321930111.

[9]. Harger JW, Meskauskas A, Dinman JD, An 'integrated model' of programmed ribosomal frameshifting, Trends Biochem. Sci 27 (2002) 448–454. 10.1016/s0968-0004(02)02149-7. [PubMed: 12217519]

[10]. Brierley I, Ribosomal frameshifting on viral RNAs, J. Gen. Virol 76 (1995) 1885–1892. 10.1099/0022-1317-76-8-1885. [PubMed: 7636469]

[11]. Lainé S, Thouard A, Komar AA, Rossignol J-M, Ribosome can resume the translation in both +1 or −1 frames after encountering an AGA cluster in Escherichia coli, Gene 412 (2008) 95–101. 10.1016/j.gene.2008.01.018. [PubMed: 18313865]

[12]. Tükenmez H, Xu H, Esberg A, Byström AS, The role of wobble uridine modifications in +1 translational frameshifting in eukaryotes, Nucleic Acids Res. 43 (2015) 9489–9499. 10.1093/nar/gkv832. [PubMed: 26283182]

[13]. Maehigashi T, Dunkle JA, Miles SJ, Dunham CM, Structural insights into +1 frameshifting promoted by expanded or modification-deficient anticodon stem loops, Proc. Natl. Acad. Sci 111 (2014) 12740–12745. 10.1073/pnas.1409436111. [PubMed: 25128388]

[14]. Bhatt PR, Scaiola A, Loughran G, Leibundgut M, Kratzel A, Meurs R, Dreos R, Gatfield D, Atkins JF, Ban N, Structural basis of ribosomal frameshifting during translation of the SARS-CoV-2 RNA genome, Science 372 (2021) 1306–1313. [PubMed: 34029205]

[15]. Zhang K, Zheludev IN, Hagey RJ, Haslecker R, Hou YJ, Kretsch R, Pintilie GD, Rangan R, Kladwang W, Li S, Wu MT-P, Pham EA, Bernardin-Souibgui C, Baric RS, Sheahan TP, D'Souza V, Glenn JS, Chiu W, Das R, Cryo-EM and antisense targeting of the 28-kDa frameshift stimulation element from the SARS-CoV-2 RNA genome, Nat. Struct. Mol. Biol 28 (2021) 747–754. 10.1038/s41594-021-00653-y. [PubMed: 34426697]

[16]. Kelly JA, Olson AN, Neupane K, Munshi S, San Emeterio J, Pollack L, Woodside MT, Dinman JD, Structural and functional conservation of the programmed −1 ribosomal frameshift signal of SARS coronavirus 2 (SARS-CoV-2), J. Biol. Chem 295 (2020) 10741–10748. 10.1074/jbc.AC120.013449. [PubMed: 32571880]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[17]. Schlick T, Zhu Q, Dey A, Jain S, Yan S, Laederach A, To Knot or Not to Knot: Multiple Conformations of the SARS-CoV-2 Frameshifting RNA Element, J. Am. Chem. Soc 143 (2021) 11404–11422. 10.1021/jacs.1c03003. [PubMed: 34283611]

[18]. Dey A, Yan S, Schlick T, Laederach A, Abolished frameshifting for predicted structure-stabilizing SARS-CoV-2 mutants: implications to alternative conformations and their statistical structural analyses, RNA 30 (2024) 1437–1450. [PubMed: 39084880]

[19]. Yan S, Zhu Q, Hohl J, Dong A, Schlick T, Evolution of coronavirus frameshifting elements: Competing stem networks explain conservation and variability, Proc. Natl. Acad. Sci 120 (2023) e2221324120. 10.1073/pnas.2221324120. [PubMed: 37155888]

[20]. Lee S, Yan S, Dey A, Laederach A, Schlick T, A Cascade of Conformational Switches in SARS-CoV-2 Frameshifting: Coregulation by Upstream and Downstream Elements, Biochemistry 64 (2025) 953–966. 10.1021/acs.biochem.4c00641. [PubMed: 39907285]

[21]. Andrews RJ, O'Leary CA, Tompkins VS, Peterson JM, Haniff HS, Williams C, Disney MD, Moss WN, A map of the SARS-CoV-2 RNA structurome, NAR Genomics Bioinforma. 3 (2021). 10.1093/nargab/lqab043.

[22]. Huston NC, Wan H, Strine MS, De Cesaris Araujo Tavares R, Wilen CB, Pyle AM, Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms, Mol. Cell 81 (2021) 584–598.e5. 10.1016/j.molcel.2020.12.041. [PubMed: 33444546]

[23]. Lan TCT, Allan MF, Malsick LE, Woo JZ, Zhu C, Zhang F, Khandwala S, Nyeo SSY, Sun Y, Guo JU, Bathe M, Näär A, Griffiths A, Rouskin S, Secondary structural ensembles of the SARS-CoV-2 RNA genome in infected cells, Nat. Commun 13 (2022) 1128. 10.1038/s41467-022-28603-2. [PubMed: 35236847]

[24]. Ziv O, Price J, Shalamova L, Kamenova T, Goodfellow I, Weber F, Miska EA, The Short- and Long-Range RNA-RNA Interactome of SARS-CoV-2, Mol. Cell 80 (2020) 1067–1077.e5. 10.1016/j.molcel.2020.11.004. [PubMed: 33259809]

[25]. Morandi E, Manfredonia I, Simon LM, Anselmi F, Van Hemert MJ, Oliviero S, Incarnato D, Genome-scale deconvolution of RNA structure ensembles, Nat. Methods 18 (2021) 249–252. 10.1038/s41592-021-01075-w. [PubMed: 33619392]

[26]. Su M-C, An atypical RNA pseudoknot stimulator and an upstream attenuation signal for −1 ribosomal frameshifting of SARS coronavirus, Nucleic Acids Res. 33 (2005) 4265–4275. 10.1093/nar/gki731. [PubMed: 16055920]

[27]. Cho C-P, Lin S-C, Chou M-Y, Hsu H-T, Chang K-Y, Regulation of Programmed Ribosomal Frameshifting by Co-Translational Refolding RNA Hairpins, PLOS ONE 8 (2013).

[28]. Pekarek L, Zimmer MM, Gribling-Burrer A-S, Buck S, Smyth R, Caliskan N, *Cis* -mediated interactions of the SARS-CoV-2 frameshift RNA alter its conformations and affect function, Nucleic Acids Res. 51 (2023) 728–743. 10.1093/nar/gkac1184. [PubMed: 36537211]

[29]. Zhu Q, Petingi L, Schlick T, RNA-As-Graphs Motif Atlas—Dual Graph Library of RNA Modules and Viral Frameshifting-Element Applications, Int. J. Mol. Sci 23 (2022) 9249. 10.3390/ijms23169249. [PubMed: 36012512]

[30]. Olson SW, Turner A-MW, Arney JW, Saleem I, Weidmann CA, Margolis DM, Weeks KM, Mustoe AM, Discovery of a large-scale, cell-state-responsive allosteric switch in the 7SK RNA using DANCE-MaP, Mol. Cell 82 (2022) 1708–1723.e10. 10.1016/j.molcel.2022.02.009. [PubMed: 35320755]

[31]. Irving PS, Weeks KM, RNAvigate: efficient exploration of RNA chemical probing datasets, Nucleic Acids Res. 52 (2024) 2231–2241. 10.1093/nar/gkae089. [PubMed: 38348910]

[32]. Hajdin CE, Bellaousov S, Huggins W, Leonard CW, Mathews DH, Weeks KM, Accurate SHAPE-directed RNA secondary structure modeling, including pseudoknots, Proc. Natl. Acad. Sci 110 (2013) 5498–5503. 10.1073/pnas.1219988110. [PubMed: 23503844]

[33]. Mustoe AM, Lama NN, Irving PS, Olson SW, Weeks KM, RNA base-pairing complexity in living cells visualized by correlated chemical probing, Proc. Natl. Acad. Sci 116 (2019) 24574–24582. 10.1073/pnas.1905491116. [PubMed: 31744869]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

[34]. Aruda J, Grote SL, Rouskin S, Untangling the pseudoknots of SARS-CoV-2: Insights into structural heterogeneity and plasticity, Curr. Opin. Struct. Biol 88 (2024) 102912. 10.1016/j.sbi.2024.102912. [PubMed: 39168046]

[35]. Leger M, Dulude D, Steinberg SV, Brakier-Gingras L, The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed −1 ribosomal frameshift, Nucleic Acids Res. 35 (2007) 5581–5592. 10.1093/nar/gkm578. [PubMed: 17704133]

[36]. Sanders CL, Curran JF, Genetic analysis of the E site during RF2 programmed frameshifting, RNA 13 (2007) 1483–1491. 10.1261/rna.638707. [PubMed: 17660276]

[37]. Rouskin S, Allan M, Aruda J, Plung J, Grote S, Martin Y, De Lajarte A, Bathe M, Discovery and Quantification of Long-Range RNA Base Pairs in Coronavirus Genomes with SEARCH-MaP and SEISMIC-RNA, (2024). 10.21203/rs.3.rs-4814547/v1.

[38]. Mikkelsen AA, Gao F, Carino E, Bera S, Simon AE, −1 Programmed ribosomal frameshifting in Class 2 umbravirus-like RNAs uses multiple long-distance interactions to shift between active and inactive structures and destabilize the frameshift stimulating element, Nucleic Acids Res. 51 (2023) 10700–10718. 10.1093/nar/gkad744. [PubMed: 37742076]

[39]. Mitchell D, Cotter J, Saleem I, Mustoe AM, Mutation signature filtering enables high-fidelity RNA structure probing at all four nucleobases with DMS, Nucleic Acids Res. 51 (2023) 8744–8757. 10.1093/nar/gkad522. [PubMed: 37334863]

[40]. Ehrhardt JE, Weeks KM, Time-Resolved, Single-Molecule, Correlated Chemical Probing of RNA, J. Am. Chem. Soc 142 (2020) 18735–18740. 10.1021/jacs.0c06221. [PubMed: 33095984]

[41]. Smola MJ, Rice GM, Busan S, Siegfried NA, Weeks KM, Selective 2′-hydroxyl acylation analyzed by primer extension and mutational profiling (SHAPE-MaP) for direct, versatile and accurate RNA structure analysis, Nat. Protoc 10 (2015) 1643–1669. 10.1038/nprot.2015.103. [PubMed: 26426499]

[42]. Busan S, Weeks KM, Accurate detection of chemical modifications in RNA by mutational profiling (MaP) with ShapeMapper 2, RNA 24 (2018) 143–148. 10.1261/rna.061945.117. [PubMed: 29114018]

[43]. Smola MJ, Calabrese JM, Weeks KM, Detection of RNA–Protein Interactions in Living Cells with SHAPE, Biochemistry 54 (2015) 6867–6875. 10.1021/acs.biochem.5b00977. [PubMed: 26544910]

[44]. Nawrocki EP, Eddy SR, Infernal 1.1: 100-fold faster RNA homology searches, Bioinformatics 29 (2013) 2933–2935. 10.1093/bioinformatics/btt509. [PubMed: 24008419]

[45]. Rivas E, RNA structure prediction using positive and negative evolutionary information, PLOS Comput. Biol 16 (2020) e1008387. 10.1371/journal.pcbi.1008387. [PubMed: 33125376]

[46]. Rivas E, Clements J, Eddy SR, Estimating the power of sequence covariation for detecting conserved RNA structure, Bioinformatics 36 (2020) 3072–3076. 10.1093/bioinformatics/btaa080. [PubMed: 32031582]

[47]. Rivas E, Clements J, Eddy SR, A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs, Nat. Methods 14 (2017) 45–48. 10.1038/nmeth.4066. [PubMed: 27819659]

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

Highlights:

- We experimentally characterize structure/function relationships for 4 coronaviruses
- Neither ensemble complexity nor individual stems predict frameshifting efficiency
- Proximal upstream sequences greatly impact frameshifting in a modular manner
- We provide tools for refining the stoichiometric expression of any two ORFs

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

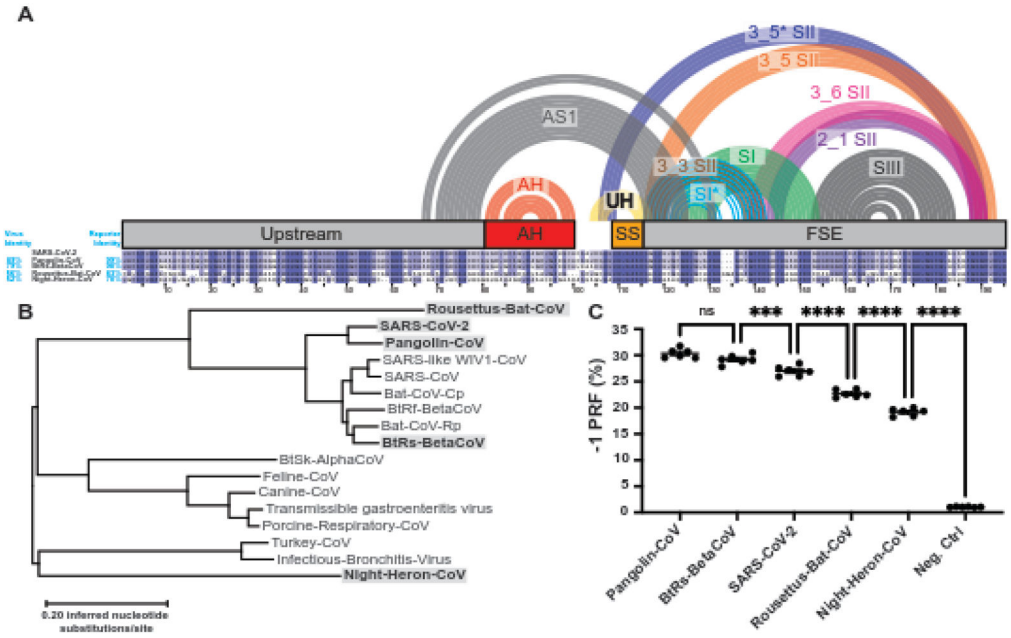


Figure 1: Viral evolution provides a platform for refining structural determinants of −1 PRF efficiency.

(A) Multisequence alignment of four coronavirus −1 PRF reporters against that of SARS-CoV-2. Colors represent sequence conservation. Percent identity for the aligned regions (reporter identity) and viral genomes (viral identity) is provided. Regions of interest are denoted above the alignment. Arcs represent previously reported substructures in the SARS-CoV-2 −1 PRF element structural ensemble. AH = Attenuator Hairpin, SS = Slippery Sequence, FSE = Frameshift Stimulation Element, AS1 = alternative stem I, UH = upstream hairpin, SI = stem I, SI* = stem I*, SII = stem II, SIII = stem III. (B) Phylogenetic tree showing the 16 coronaviruses most similar to SARS-CoV-2 by co-variation analysis reported in Schlick et. al., 2021 [17]. Branch length indicates evolutionary time. Viruses compared in this study are bolded and highlighted in gray. (C) Results of −1 PRF luciferase assays for each reporter. N = 6 biological replicates each averaged from 3 technical replicates. Selected results from Tukey's multiple comparisons shown. ns = not significant, **p<0.01, ****p<0.0001.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

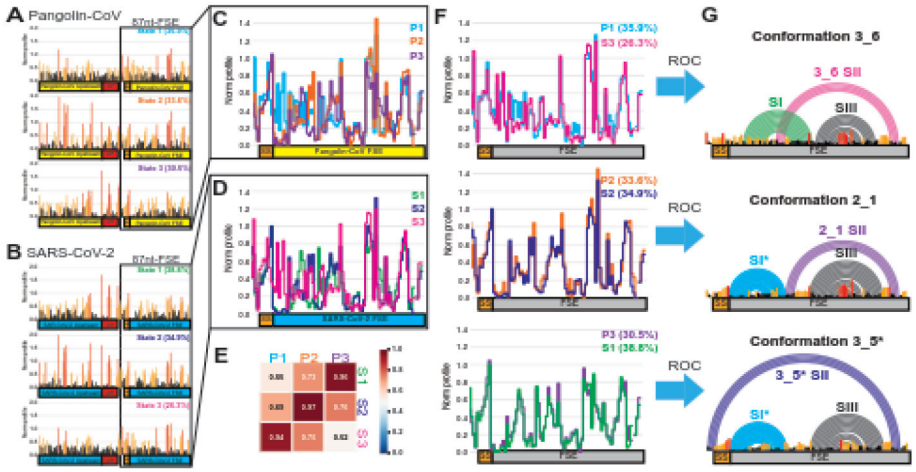


Figure 2. A modeling-free framework for comparing structures in evolutionarily conserved elements.

(**A,B**) DMS reactivity profiles for each conformational state identified by Dance-MaP analysis of Pangolin-CoV and SARS-CoV-2. Population percentages for each state are shown in parentheses. Red = high reactivity, orange = moderate reactivity, black = low reactivity. (**C,D**) Skyline plots of overlapping DMS reactivity profiles for each Pangolin-CoV or SARS-CoV-2 state respectively over the 87nt-FSE region. (**E**) Heatmap depicting each pairwise correlation coefficient between DMS reactivities profiles for each Pangolin-CoV and SARS-CoV-2 state for the 87nt-FSE region. R-values are listed on each tile. (**F**) Skyline plots of overlapping DMS reactivity profiles for each set of highly correlated Pangolin-CoV and SARS-CoV-2 states in panel E. Population percentages for each state are shown in parentheses. (**G**) Arc diagrams depicting the previously reported [18] SARS-CoV-2 secondary structure models most consistent with the DMS reactivity profiles for each combined state as determined by Receiver Operating Characteristic (ROC) analysis. The DMS reactivity profile for each previously uncharacterized Pangolin-CoV state is projected over each respective structure model. Red = high reactivity, orange = moderate reactivity, black = low reactivity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

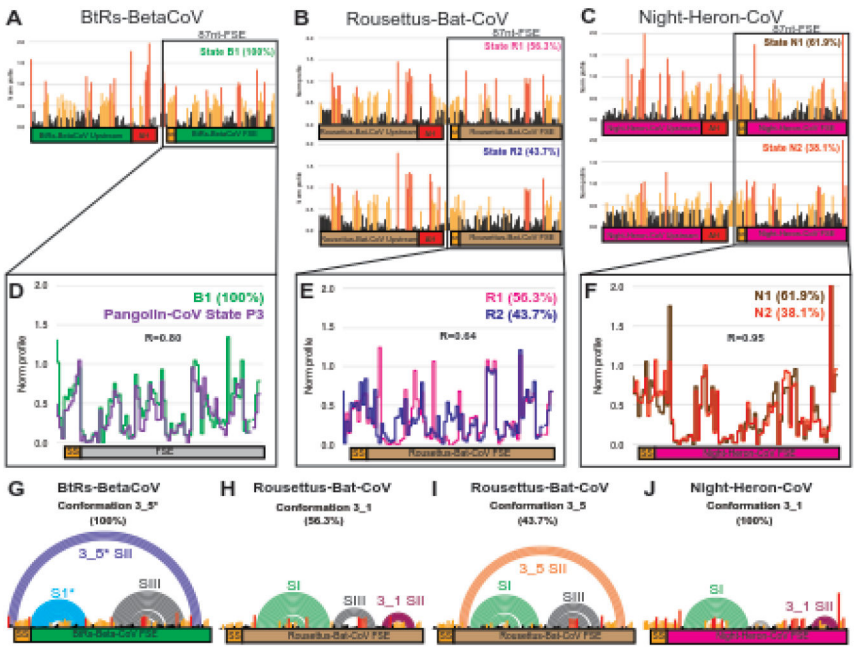


Figure 3. Ensemble complexity is not the main determinant of −1 PRF efficiency, and coronavirus FSEs converge on a limited set of substructures.

(**A-C**) DMS reactivity profiles for each state identified by Dance-MaP analysis for BtRs-BetaCoV, Rousettus-Bat-CoV, and Night-Heron-CoV respectively. Population percentages for each state are shown in parentheses. Red = high reactivity, orange = moderate reactivity, black = low reactivity. (**D**) Skyline plots of the DMS reactivity profile for BtRs-BetaCoV in the 87nt-FSE region overlaid with that of Pangolin-CoV state P3. Correlation coefficient R=0.80. (**E-F**) Skyline plots of the DMS reactivity profiles for each Rousettus-Bat-CoV or Night-Heron-CoV state respectively in the 87nt-FSE region. Correlation coefficients (R-values) between states are listed on each plot. (**G**) Arc diagram of the previously described 3_5* conformation assigned to BtRs-BetaCoV through ROC analysis. BtRs-BetaCoV DMS reactivities are overlaid on the structure model. Red = high reactivity, orange = moderate reactivity, black = low reactivity. (**H-J**) Arc diagrams of the secondary structure models for each Rousettus-Bat-CoV and Night-Heron-CoV 87nt-FSE conformation respectively, obtained using ShapeKnots. The percentages of each conformation are shown in parentheses. DMS reactivities for each state are overlaid on the structure models (the averaged reactivity for both states is shown for Night-Heron-CoV). Red = high reactivity, orange = moderate reactivity, black = low reactivity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

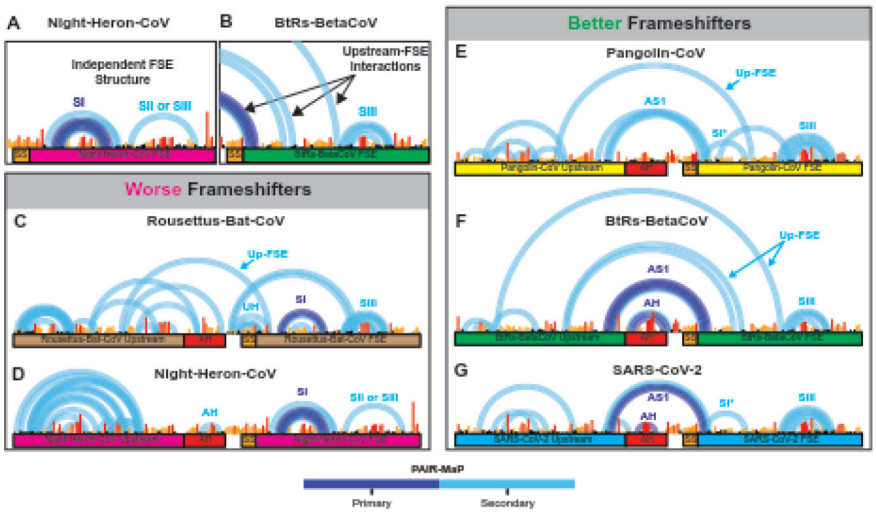


Figure 4. Correlated chemical probing suggests AS1 separates better frameshifters from worse ones.

(**A-B**) Arc diagrams depicting PAIRs interactions for Night-Heron-CoV and BtRs-Beta-CoV respectively for the 87nt-FSE region. (**C-D**) PAIRs interactions across the entire viral reporter element for the worse frameshifting Rousettus-Bat-CoV and Night-Heron-CoV reporters respectively. (**E-G**) PAIRs interactions across the entire viral reporter element for the better frameshifting Pangolin-CoV, BtRs-BetaCoV, and SARS-CoV-2 reporters respectively. DMS reactivity profiles for each reporter are overlaid on their respective plots. Red = high reactivity, orange = moderate reactivity, black = low reactivity.

Author Manuscript Author Manuscript Author Manuscript Author Manuscript

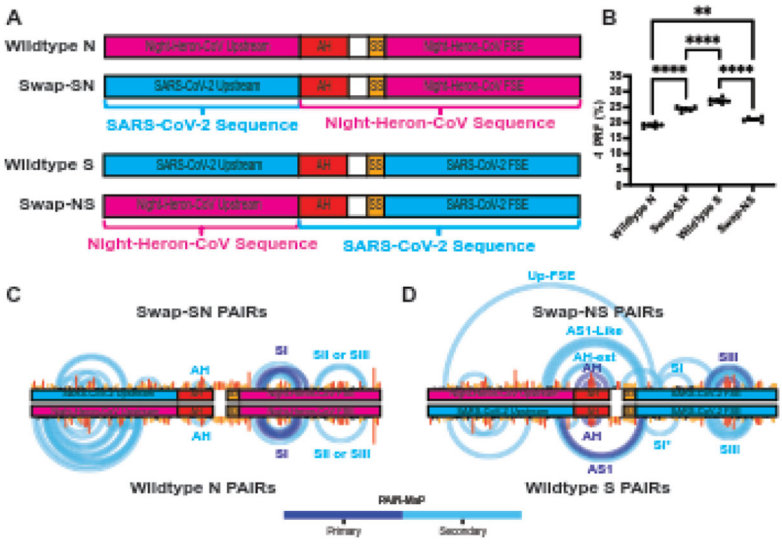


Figure 5. Sequences immediately upstream of the AH contain strong regulatory potential independent of AS1 formation.

(**A**) Schematic depiction of each swap mutant design compared to the respective wildtype reporters. The sequences upstream of the AH were swapped with those of the indicated virus. (**B**) Results of the −1 PRF luciferase assays for each reporter. N = 6 biological replicates each averaged from 3 technical replicates. Selected results from Tukey's multiple comparison shown. Values for wildtype reporters reproduced from Figure 1 (see methods). ns = not significant, **p<0.01, ***p<0.0001. (**C-D**) Arc diagrams depicting PAIRs interactions for each swap mutant (top) compared with their respective wildtype reporter (bottom). DMS reactivity profiles for each reporter are overlaid on their respective plots. Red = high reactivity, orange = moderate reactivity, black = low reactivity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1: Comparing 87nt-FSE ensembles and model-supported substructures to functional outcomes in −1 PRF efficiency.

				Amount of Structure in Ensemble (%)							
		# States					SII variant				
Reporter Construct	**−1 PRF**	**Total**	**FSE**	**SI**	**SI***	**SIII**	**3_6**	**3_5**	**3_5***	**3_1**	**2_1**
Pangolin-CoV	30%	3	3	36	64	100	36	-	31	-	34
BtRs-Beta-CoV	29%	1	1	-	100	100	-	-	100	-	-
SARS-CoV-2	27%	3	3	26	74	100	26	-	39	-	35
Rousettus-Bat-CoV	23%	2	2	100	-	100	-	44	-	56	-
Night-Heron-CoV	19%	2	1	100	-	-	-	-	-	100	-

Reporters are listed in decreasing order by −1 PRF efficiency. The number of DANCE-MaP states across the entire viral insert (total) and 87nt-FSE region (FSE) are listed. The relative percentage of each major substructure within the structural ensemble of each reporter is listed. Dashes indicate a substructure does not appear in the ensemble.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2: Comparing ensembles and PAIRs-supported substructures to functional outcomes in −1 PRF efficiency.

				# States		PAIRs-Supported Structures					
Reporter Construct	**FSE**	**−1 PRF**	**PRF**	**Total**	**FSE**	**AS1**	**SI**	**SI***	**AH**	**SIII**	**Up-FSE**
Night-Heron-CoV	NH	19%	-	2	1	**N**	**Y**	**N**	**Y**	**Y**	**N**
Swap-SN	NH	24%	**+5%**	3	2	**N**	**Y**	**N**	**Y**	**Y**	**N**
SARS-CoV-2	SARS	27%	-	3	3	**Y**	**N**	**Y**	**Y**	**Y**	**N**
Swap-NS	SARS	21%	**−6%**	3	2	**N**	**Y**	**N**	**Y**	**Y**	**Y**

Reporters are separated by FSE region. Calculated −1 PRF efficiencies are shown for each reporter as well as change in −1 PRF efficiency for mutant reporters compared to the respective wildtype reporter for each FSE. The number of DANCE-MaP states across the entire viral insert (total) and 87nt-FSE region (FSE) are listed. Substructures supported by PAIRs are indicated with a Y. Structures not supported by PAIRs are indicated with an N. Up-FSE represents any upstream-FSE interaction other than AS1.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3. Coronavirus reporter viral insert sequences.

Reporter	Length	
SARS-CoV-2	192	TACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCATGCTTCAGTCA
Pangolin-CoV	192	TACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCATGCTTCAGTCA
BtRs-BetaCoV	192	TACACTTAGGAACACAGTCTGTACCGTCTGCGGGATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCATGATGCAGTCT
Rousettus-Bat-CoV	189	CATACGTAACGTGATTTGTTCCGTGTGCAACATGTGGCAAGGATATGGCTGTCCATGTGCTAGCTTACGTGAAGTGAGCCTGCACGCTAAT
Night-Heron-CoV	189	TGCACTTACACATGAACCTTGCACTGCTTGTCACAGGTGGGTAAATCATGATTGTACGTGTACAAACCTTCAATTGGGTTATAGTTCAAACC
Swap-SN mutant	195	TACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCATGGGTTATAGT
Swap-NS mutant	189	TGCACTTACACATGAACCTTGCACTGCTTGTCACAGGTGGGTAAATCATGATTGTACGTGTACAAACCTTCAATTGATGCTTCAGTCAGCTC
Negative Ctrl	193	TACACTTAAAAACACAGTCTGTACCGTCTGCGGTATGTGGAAAGGTTATGGCTGTAGTTGTGATCAACTCCGCGAACCCAGGTTATAGTTC

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 4. Primers used to prepare next-generation sequencing libraries.

Primer	Sequence
FW	5’-GACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNNNNCCTCGAGAGAAGCCAGACAC-3’
RV	5’-CCCTACACGACGCTCTTCCGATCTNNNNNGGTGGCAGATCCGAAAGGTCC-3’