Structural and functional conservation of the programmed -1 ribosomal frameshift signal of SARS-CoV-2

Jamie A. Kelly and Jonathan D. Dinman*

Department of Cell Biology and Molecular Genetics, University of Maryland, College Park MD 20742

*Corresponding author: Jonathan D. Dinman

E-mail: dinman@umd.edu

Running title: Frameshifting in SARS-CoV-2

Keywords: COVID-19, coronavirus, programmed -1 ribosomal frameshifting (-1 PRF), translation, RNA, structure, virus, (+) ssRNA

Abstract.

17 years after the SARS-CoV epidemic, the world is facing the COVID-19 pandemic. COVID-19 is caused by a coronavirus named SARS-CoV-2. Given the most optimistic projections estimating that it will take more than a year to develop a vaccine, our best short term strategy may lie in identifying virus-specific targets for small molecule interventions. All coronaviruses utilize a molecular mechanism called -1 PRF to control the relative expression of their proteins. Prior analyses of SARS-CoV revealed that it utilizes a structurally unique three-stemmed mRNA pseudoknot to stimulate high rates of -1 PRF, that it also harbors a -1 PRF attenuation element. Altering -1 PRF activity negatively impacts virus replication, suggesting that this molecular mechanism may be therapeutically targeted. Here we present a comparative analysis of the original SARS-CoV and SARS-CoV-2 frameshift signals. Structural analyses reveal that the core -1 PRF signal, composed of the U UUA AAC slippery site and three-stemmed mRNA pseudoknot is highly conserved. In contrast, the upstream attenuator hairpin is less well conserved. Functional assays revealed that both elements promote similar rates of -1 PRF and that silent coding mutations in the slippery site strongly ablate -1 PRF activity. We suggest that molecules that were previously identified as inhibiting SARS-CoV mediated -1 PRF may serve as lead compounds to counter the current pandemic.

Introduction

SARS-CoV2, the etiological agent of COVID-19, is a member of the coronavirus family (1). Coronaviruses have (+) ssRNA genomes that harbor two long open reading frames (ORF) which occupy the 5' ~ two-thirds of the genomic RNA (ORF1 and ORF2), followed by several ORFs that are expressed late in the viral replication cycle from subgenomic RNAs (sgRNAs) (**Fig. 1A**) (2). In general the immediate early proteins encoded by ORF1 are involved in ablating the host cellular innate immune response, while the early proteins encoded in ORF2 are involved in genome replication and RNA synthesis. These functions include generation the minus-strand replicative intermediate, new plus-strand genomic RNAs, and subgenomic RNAs which mostly encode structural, late proteins. ORF2 is out of frame with respect to ORF1, and all coronaviruses utilize a molecular mechanism called programmed -1 ribosomal frameshifting (-1 PRF) as a means to synthesize the ORF2 encoded proteins (3, 4). -1 PRF is a mechanism in which *cis*-acting elements in the mRNA direct elongating ribosomes to shift reading frame by 1 base in the 5' direction [reviewed in (3, 4). The use of a -1 PRF mechanism for expression of a viral gene was first published in 1985 for the Rous sarcoma virus (5). A -1 PRF mechanism was shown to be required to translate ORF1ab in a coronavirus, Avain Infectious Bronchitis Virus (IBV), two years later (6). In coronaviruses -1

PRF functions as a developmental switch, and mutations and small molecules that alter this process have deleterious effects on virus replication (7, 8).

The –1 PRF signal can be broken down into three discrete parts: the "slippery site", a linker region, and a downstream stimulatory region of mRNA secondary structure, typically an mRNA pseudoknot [reviewed in (3)]. The primary sequence of the slippery site and its placement in relation to the incoming translational reading frame is critical: it must be N NNW WWZ (codons are shown in the incoming or 0-frame), where N is a stretch of three identical nucleotides, W is either AAA or UUU, and Z ≠ G. The linker region is less well-defined, but typically is short (1 – 12 nt long) and is thought to be important for determining the extent of –1 PRF in a virus-specific manner. The function of the downstream secondary structure is to induce elongating ribosomes to pause, a critical step for efficient –1 PRF to occur [reviewed in (9)]. The generally accepted mechanism of –1 PRF is that the mRNA secondary structure directs elongating ribosomes to pause with its A- and P-site bound aminoacyl- (aa-) and peptidyl-tRNAs are positioned over the slippery site. The sequence of the slippery site allows for re-pairing of the tRNAs to the –1 frame codons after they "simultaneously slip" by one base in the 5' direction along the mRNA. The subsequent resolution of the downstream mRNA secondary structure allows the ribosome to continue elongation of the nascent polypeptide in the new translational reading frame. The downstream stimulatory elements are most commonly H-type mRNA pseudoknots, so called because they are composed of two co-axially stacked stem-loops where the second stem is formed by base pairing between sequence in the loop of the first-stem loop, and additional downstream sequence (10). The SARS-CoV pseudoknot is more complex because it contains a third, internal stem-loop element (11–13). Mutations affecting this structure decreased rates of -1 PRF, and had deleterious effects on virus propagation, thus suggesting that it may present a target for small molecule therapeutics (7, 8). In addition, the presence of a hairpin located immediately 5' of the slippery site has been reported to regulate -1 PRF by attenuating its activity (14). Here, we report on the -1 PRF signal from SARS-CoV2. The core -1 PRF signal is nearly identical to the SARS-CoV -1 PRF signal, containing only a single nucleotide difference, a C to A. This change maps to a loop region in the molecule that is not predicted to impact on the structure of the three-stemmed pseudoknot. The attenuator hairpin appears to be less-well conserved. However, functional analyses reveal that both elements promote nearly identical rates of -1 PRF, suggesting that this element has been structurally and functionally conserved.

Results and Discussion

Comparative structural analyses of the two -1 PRF signals. The core of the SARS-CoV -1 PRF signal begins with the U UUA AAC slippery site. This is followed by a 6 nt spacer region, which is then followed by the three-stemmed -1 PRF stimulating mRNA pseudoknot. A second regulatory element, called the attenuator hairpin, is located 5' of the slippery site. Pairwise analysis of the SARS-CoV and SARS-CoV-2 frameshift signals revealed that the sequence of the attenuator hairpin was less well conserved than the frameshift-stimulating pseudoknot (**Fig. 1B**). The structure of the SARS-CoV -1 PRF signal was previously determined to include a three-stemmed mRNA pseudoknot (11). Using this as a guide, the single C to A base difference between the core SARS-CoV and SARS-CoV-2 -1 PRF signals maps to a loop that is not predicted to alter the structure of the -1 PRF stimulating element (7) (**Fig. 1C**). In contrast, the attenuator hairpin contains six differences in the nucleotide sequence between the two viruses, and the SARS-CoV-2 element is predicted to be less stable than its SARS-CoV counterpart (**Fig. 1D**). This suggests that the attenuation activity of this element is not as important as the ability to direct -1 PRF *per se*.

Comparative functional analyses of the two -1 PRF signals. Standard dual-luciferase assays were used to monitor -1 PRF activities of the two -1 PRF signals (15, 16) in HEK and HeLa cells. For both of the elements, -1 PRF activity was ~20% in HEK (**Fig. 2A**) and ~30% in HeLa (**Fig. 2B**). Amino acid sequence silent coding mutation of the U UUA AAC slippery sites to C CUC AAC (the incoming 0-frame codons are indicated by spaces) ablated -1 PRF activity in both cases, to less than 1% (**Fig. 2**). These findings support the hypothesis that structure and function of the -1 PRF signals has been conserved between the two viruses. Unfortunately, the dual luciferase reporter pJD2359 which was used to generate the other reporters used in this study does not contain the attenuator sequence. Thus, differences in this element were not able to be assayed.

A potential target for antiviral interventions. Mutations that alter rates of -1 PRF efficiency have been shown to be deleterious to replication of SARS-CoV (7, 8). A subsequent study described a series of antisense peptide nucleic acids that inhibited SARS-CoV mediated -1 PRF and virus replication (17), and mRNA pseudoknot dimerization was subsequently shown to be important for -1 PRF activity (18). Separately, a high throughput screening approach was used to identify a small molecule that inhibited SARS-CoV mediated -1 PRF (19). Interestingly, this was shown to limit the conformational plasticity of the -1 PRF stimulating pseudoknot, suggesting that targeting the conformational dynamics of pseudoknots may be an effective strategy for anti-viral drug design (20). Given the >1 year

best-case scenario timeframe for development of a vaccine, similar approaches targeting -1 PRF may be of interest for controlling the SARS-CoV-2 pandemic.

Experimental Procedures

Identification of the SARS-CoV2 -1 PRF signal and computational methods. The SARS-CoV-2 -1 PRF signal was identified from the complete genome sequence (NCBI Ref Seq NC_045512.2). The EMBOSS Water pairwise alignment tool was used to identify sequences in the SARS-CoV-2 genome most similar to the SARS-CoV -1 PRF sequence. One hit was reported between bases 13461 and 13547 of SARS-CoV-2 that was 98.9% identical to the original SARS sequence. The SARS-CoV-2 sequence contains a single point mutation from C to A at base 13533. EMBOSS Water was used to generate pairwise alignments between sequences derived from SARS-CoV (GenBank entry NC_004718.3 begin nt 13361, end nt 13478) and SARS-CoV-2 (Genbank entry NC_045512.2, begin nt 13431, end nt 13547).

Plasmid construction and bacterial transformation. Reporter plasmids for SARS-CoV-2 were generated by oligonucleotide site directed mutagenesis using the NEB Q5 Site directed mutagenesis kit (cat. #E0554S). A single C to A point mutation was introduced into pJD2359 (SARS-CoV pSGDluc reporter plasmid) (8) at base 1873 corresponding to the point mutation in the SARS-CoV-2 genome. Site directed mutagenesis primers were synthesized and purified by IDT. Products were transformed in to DH5α Escherichia coli cells (NEB) and spread onto LB agar plates containing 50μg/mL carbenicillin. Positive clones were verified by DNA sequencing (Genewiz). Additional frameshift reporter negative controls were constructed using site-directed mutagenesis. Silent mutations disrupting the -1 PRF slippery site from UUU UUA AAC to UUC CUC AAC (named ssM) were made using the NEB site directed mutagenesis kit.

Cell culture. Human embryonic kidney (HEK293T/17) (CRL-11268) and HeLa (CCL-2) cells were purchased from the American Type Culture Collection (Manassas, VA). HEK293T cells were maintained in Dulbecco's modified Eagle medium (DMEM) (Fisher Scientific 10-013-CV) supplemented with 10% fetal bovine serum (FBS) (Fisher Scientific 26140-079), 1% GlutaMAX (35050-061), 1% nonessential amino acids (NEAA) (Fisher Scientific 11140-050), 1% HEPES buffer (Fisher Scientific 15630-030) and 1x Penicillin-streptomycin (Fisher Scientific 15140-122) at 37°C in 5% CO_2. HeLa cells were maintained in DMEM supplemented with 10% FBS, 1% GlutaMAX and 1x Penicillin-streptomycin at 37°C in 5% CO_2.

Plasmid transfection. HEK293T and HeLa cells were seeded at 4 x 10^4 cells per well into 24-well plates. Cells were transfected 24 hours after seeding with 500ng dual luciferase reporter

plasmid using Lipofectamine3000 (Invitrogen L3000015) per manufacturer's protocol.

Dual luciferase assays and calculation of -1 frameshifting efficiency. Frameshifting efficiency of the reporter plasmids were assayed as previously described (15, 16) using a dual luciferase reporter assay system kit (Promega). 24 hours post transfection, cells were washed with 1x phosphate-buffered saline (PBS) then lysed with 1x passive lysis buffer (E194A, Promega). Cell lysates were assayed in triplicate in a 96-well plate and luciferase activity was quantified using a GloMax microplate luminometer (Promega). Percent frameshift was calculated by averaging the three firefly or *Renilla* luciferase technical replicate reads per sample then forming a ratio of firefly to *Renilla* luminescence per sample. Each sample luminescence ratio was compared to a readthrough control set to 100%. The ratio of ratios for each sample is the percent frameshift for the sample. A minimum of three biological replicates were assayed for each sample, each of which were assayed in triplicate.

Data Availability. Full datasets of -1 PRF assays are available upon request.

Acknowledgements

We wish to thank Kevin Tu for his help with reagent preparation.

Funding: This work was supported by Defense Threat Reduction Agency Grant HDTRA1-13-1-0005; NIGMS, National Institutes of Health (NIH) Grant R01 GM117177 (to J.D.D.), and NIH Institutional Training Grant 2T32AI051967-06A1 (to J. A. K.). "The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest. The authors declare that they have no conflicts of interest with the contents of this article.

References

1. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2 (2020) *Nat. Microbiol.* 10.1038/s41564-020-0695-z
2. Brian, D. A., and Baric, R. S. (2005) Coronavirus Genome Structure and Replication. in *Coronavirus Replication and Reverse Genetics*, pp. 1–30, Springer-Verlag, Berlin/Heidelberg, **287**, 1–30
3. Dinman, J. D. (2012) Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip.Rev.RNA.* **3**, 661–673
4. Atkins, J. F., Loughran, G., Bhatt, P. R., Firth, A. E., and Baranov, P. V (2016) Ribosomal frameshifting and transcriptional slippage: From genetic steganography and cryptography to adventitious use. *Nucleic Acids Res.* **44**, 7007–7078
5. Jacks, T., and Varmus, H. E. (1985) Expression of the Rous Sarcoma Virus pol gene by ribosomal frameshifting. *Science (80-.).* **230**, 1237–1242
6. Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D., and Inglis, S. C. (1987) An efficient ribosomal frame-shifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* **6**, 3779–3785
7. Plant, E. P. P., Rakauskaite, R., Taylor, D. R. R., and Dinman, J. D. D. (2010) Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J.Virol.* **84**, 4330–4340
8. Plant, E. P., Sims, A. C., Baric, R. S., Dinman, J. D., and Taylor, D. R. (2013) Altering SARS coronavirus frameshift efficiency affects genomic and subgenomic RNA production. *Viruses*. **5**, 279–94
9. Rodnina, M. V, Korniy, N., Klimova, M., Karki, P., Peng, B.-Z., Senyushkina, T., Belardinelli, R., Maracci, C., Wohlgemuth, I., Samatova, E., and Peske, F. (2019) Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res.* 10.1093/nar/gkz783
10. Puglisi, J. D., Wyatt, J. R., and Tinoco Jr., I. (1988) A pseudoknotted RNA oligonucleotide. *Nature*. **331**, 283–286
11. Plant, E. P., Perez-Alvarado, G. C., Jacobs, J. L., Mukhopadhyay, B., Hennig, M., and Dinman, J. D. (2005) A Three-Stemmed mRNA Pseudoknot in the SARS Coronavirus Frameshift Signal. *PLoS.Biol.* **3**, 1012–1023
12. Ramos, F. D., Carrasco, M., Doyle, T., and Brierley, I. (2004) Programmed -1 ribosomal frameshifting in the SARS coronavirus. *Biochem.Soc.Trans.* **32**, 1081–1083
13. Baranov, P. V, Henderson, C. M., Anderson, C. B., Gesteland, R. F., Atkins, J. F., and Howard, M. T. (2005) Programmed ribosomal frameshifting in decoding the SARS-CoV genome. *Virology*. **332**, 498–

510

14. Cho, C.-P., Lin, S.-C., Chou, M.-Y., Hsu, H.-T., and Chang, K.-Y. (2013) Regulation of programmed ribosomal frameshifting by co-translational refolding RNA hairpins. *PLoS One*. **8**, e62283
15. Harger, J. W., and Dinman, J. D. (2003) An in vivo dual-luciferase assay system for studying translational recoding in the yeast Saccharomyces cerevisiae. *RNA*. **9**, 1019–1024
16. Jacobs, J. L., and Dinman, J. D. (2004) Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Res.* **32**, e160
17. Ahn, D. G., Lee, W., Choi, J. K., Kim, S. J., Plant, E. P., Almazan, F., Taylor, D. R., Enjuanes, L., and Oh, J. W. (2011) Interference of ribosomal frameshifting by antisense peptide nucleic acids suppresses SARS coronavirus replication. *Antivir. Res.* **91**, 1–10
18. Ishimaru, D., Plant, E. P., Sims, A. C., Yount, B. L., Roth, B. M., Eldho, N. V, Pérez-Alvarado, G. C., Armbruster, D. W., Baric, R. S., Dinman, J. D., Taylor, D. R., Hennig, M., and Hennig, M. (2013) RNA dimerization plays a role in ribosomal frameshifting of the SARS coronavirus. *Nucleic Acids Res.* **41**, 2594–608
19. Park, S. J., Kim, Y. G., and Park, H. J. (2011) Identification of RNA pseudoknot-binding ligand that inhibits the -1 ribosomal frameshifting of SARS-coronavirus by structure-based virtual screening. *J.Am.Chem.Soc.* **133**, 10094–10100
20. Ritchie, D. B., Soong, J., Sikkema, W. K. A., and Woodside, M. T. (2014) Anti-frameshifting Ligand Reduces the Conformational Plasticity of the SARS Virus Pseudoknot. *J. Am. Chem. Soc.* **136**, 2196–2199

Figure Legends

Figure 1. Structural comparison of the SARS-CoV and SARS-CoV-2 -1 PRF signals. A. Car**B.** Pairwise analysis of the two -1 PRF signals. The attenuator elements and three-stemmed pseudoknot sequences are boxed as indicated. The U UUA AAC slippery site is underlined. **C.** Structure of the SARS-CoV -1 PRF signal (11) is composed of the 5' slippery site, a 6 nt spacer, and the three-stemmed pseudoknot stimulatory element. The single base difference in SARS-CoV-2 (red) maps to the short loop linking Stems 2 and 3. **D.** Comparison of the SARS-CoV and SARS-CoV-2 -1 PRF attenuator elements. SARS-CoV-2 specific bases are indicated in red.

Figure 2. Functional characterization of the SARS-CoV and SARS-CoV-2 -1 PRF signals. The efficiencies of -1 PRF promoted by the wild-type (U UUA AAC) and silent slippery site mutant (C CUC AAC) -1 PRF signals were assayed in HEK cells using dual-luciferase assays as previously described (15, 16). Assays were performed as three technical replicates, and were repeated a minimum three times apiece (biological replicates). **A.** Assays performed in HEK cells. **B.** Assays performed in HeLa cells. Error bars denote S.E.M.

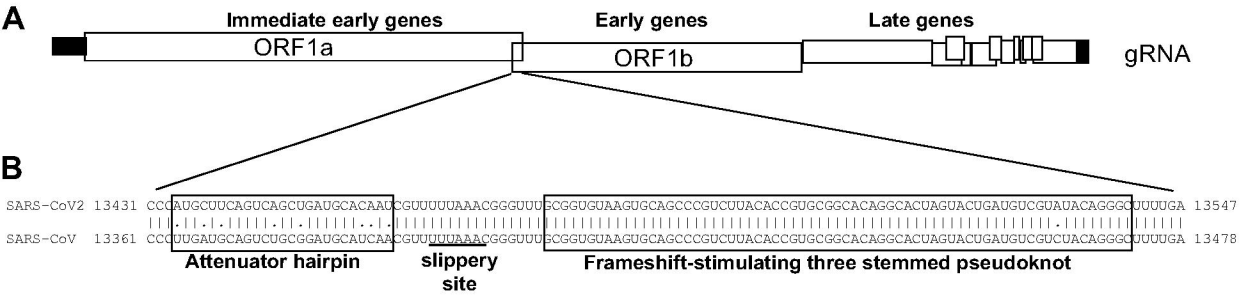

Figure 1

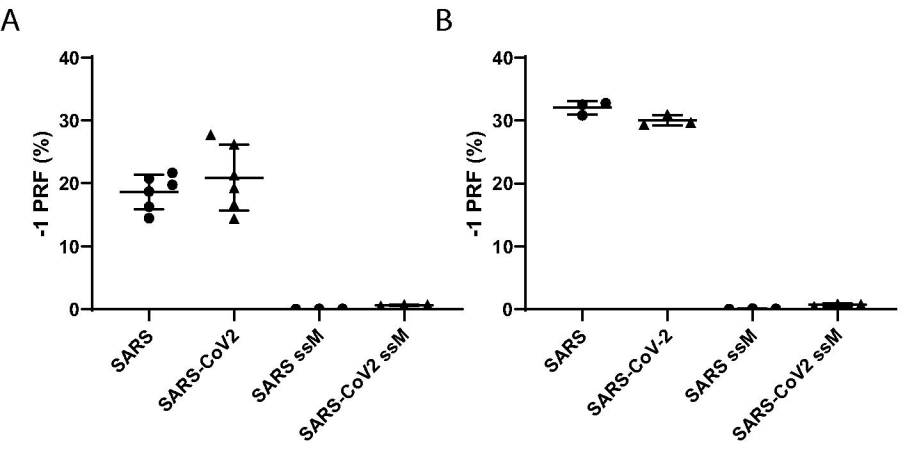

Figure 2