An RNA pseudoknot stimulates HTLV-1 *pro-pol* programmed −1 ribosomal frameshifting

ELIZA THULSON,[1,5] ERIK W. HARTWICK,[2,3,5] ANDREW COOPER-SANSONE,[1] MARCUS A.C. WILLIAMS,[1] MARY E. SOLIMAN,[4] LEILA K. ROBINSON,[4] JEFFREY S. KIEFT,[2,3] and KATHRYN D. MOUZAKIS[4]

[1]Department of Chemistry and Biochemistry, Fort Lewis College, Durango, Colorado 81301, USA

[2]Department of Biochemistry and Molecular Genetics, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA

[3]RNA Bioscience Initiative, University of Colorado Denver School of Medicine, Aurora, Colorado 80045, USA

[4]Department of Chemistry and Biochemistry, Loyola Marymount University, Los Angeles, California 90045, USA

ABSTRACT

Programmed −1 ribosomal frameshifts (−1 PRFs) are commonly used by viruses to regulate their enzymatic and structural protein levels. Human T-cell leukemia virus type 1 (HTLV-1) is a carcinogenic retrovirus that uses two independent −1 PRFs to express viral enzymes critical to establishing new HTLV-1 infections. How the *cis*-acting RNA elements in this viral transcript function to induce frameshifting is unknown. The objective of this work was to conclusively define the 3′ boundary of and the RNA elements within the HTLV-1 *pro-pol* frameshift site. We hypothesized that the frameshift site structure was a pseudoknot and that its 3′ boundary would be defined by the pseudoknot's 3′ end. To test these hypotheses, the in vitro frameshift efficiencies of three HTLV-1 *pro-pol* frameshift sites with different 3′ boundaries were quantified. The results indicated that nucleotides included in the longest construct were essential to highly efficient frameshift stimulation. Interestingly, only this construct could form the putative frameshift site pseudoknot. Next, the secondary structure of this frameshift site was determined. The dominant structure was an H-type pseudoknot which, together with the slippery sequence, stimulated frameshifting to 19.4(±0.3)%. The pseudoknot's critical role in frameshift stimulation was directly revealed by examining the impact of structural changes on HTLV-1 *pro-pol* −1 PRF. As predicted, mutations that occluded pseudoknot formation drastically reduced the frameshift efficiency. These results are significant because they demonstrate that a pseudoknot is important to HTLV-1 *pro-pol* −1 PRF and define the frameshift site's 3′ boundary.

Keywords: HTLV-1; frameshifting; RNA pseudoknot; chemical probing; translation; retrovirus

INTRODUCTION

A programmed ribosomal frameshift (PRF) is a type of translational reprogramming event that changes the ribosome's reading frame during elongation. These events allow ribosomes access to alternative reading frames by inducing a −2, −1, +1, or +2 nt ribosomal slip on the transcript. −1 PRFs are used by many positive-strand RNA viruses (e.g., retroviruses like human immunodeficiency virus type 1 [HIV-1] and human T-cell leukemia virus type 1 [HTLV-1]; coronaviruses like severe acute respiratory syndrome [SARS]) as well as cellular mRNAs, to direct the ribosome into a −1 reading frame (Dinman 2012a,b; Caliskan et al. 2015; Korniy et al. 2019; Rodnina et al. 2019). The purpose of a −1 PRF differs for viruses and cellular RNAs. For viruses, a −1 PRF allows the ribosome access to alternate reading frames that code for viral enzymes (Giedroc

and Cornish 2009; Caliskan et al. 2015). For cellular mRNAs, −1 PRF events appear to trigger nonsense-mediated mRNA decay (NMD), which influences mRNA stability (Dinman 2012a,b).

−1 PRF events are stimulated by *cis*-acting RNA elements within a transcript. Each −1 PRF occurs at a frameshift site that includes a 7-nt "slippery" sequence, a spacer sequence, and a downstream structure (Brierley 1995; Farabaugh 1996). The slippery sequence (N-NNW-WWH in the 0 frame, IUPAC notation) is indispensable and alone can increase the basal rate of ribosomal frameshifting from <0.005% per codon (Kurland 1992; Stahl et al. 2002) to as high as ∼1% (Brierley et al. 1992; Giedroc et al. 2000; Kim et al. 2001; Yu et al. 2010). Here, "N" can be any nt (A, U, C, or G), "W" can be A or U, and "H" can be any nt except for

G. This sequence is slippery because when it is positioned within the ribosomal A- and P-sites and the ribosome slips into the −1 reading frame, a thermodynamically allowable set of base pairs remains between the mRNA codons (now in the −1 frame) and the tRNA anticodons (Demeshkina et al. 2012). The N-NNW-WWH pattern ensures that the only differences in base-pairing in the alternate reading frame will occur in the wobble position of the codons (0 frame codons: NNW [P-site], WWH [A-site]; −1 frame codons: NNN [P-site], WWW [A-site]). The spacer region, usually 5–8 nt in length, separates the slippery sequence from the downstream structure (Dinman 2012b). This structure, typically a very stable stem–loop or pseudoknot, acts as a steric block to translation (Plant 2003; Giedroc and Cornish 2009; Tholstrup et al. 2012). The structure's resistance to unwinding promotes ribosomal stalling over the slippery sequence and modifies a step within elongation, EF-G-catalyzed translocation, in a way that promotes ribosomal frameshifting (Caliskan et al. 2014, 2015; Kim and Tinoco 2017; Wu et al. 2018). Together, these RNA elements direct the ribosome into the −1 reading frame at a defined frequency, or frameshift efficiency (Reil et al. 1993; Brierley et al. 2010; Atkins et al. 2016).

−1 PRFs are commonly used by viruses to regulate viral enzymatic and structural protein levels (Plant et al. 2010; Tholstrup et al. 2012; Huang et al. 2013). In addition to increasing the RNA coding capacity, gene placement in alternate reading frames provides a point of regulatory control for viruses. Many viruses require higher concentrations of their internal structural proteins than their enzymes when assembling infectious virus particles (Plant et al. 2010; Dinman 2012b). The *gag* and *pol* open reading frames usually encode these structural and enzymatic polyproteins, respectively. To achieve a precise protein ratio, *pol* is often positioned downstream and in the −1 reading frame relative to *gag*. Therefore, the frameshift efficiency controls the molar ratio of viral structural to enzymatic proteins produced (Giedroc et al. 2000). This ratio appears to be critical for the successful replication of many viruses (Plant et al. 2010; Dinman 2012b), including the SARS-associated coronavirus (Plant et al. 2010) and the HIV-1 retrovirus (Park and Morrow 1991; Karacostas et al. 1993; Hung et al. 1998; Biswas et al. 2004; Garcia-Miranda et al. 2016). Whether this ratio is similarly critical for HTLV-1 replication is unknown.

HTLV-1 is a human carcinogenic retrovirus (Tagaya and Gallo 2017; Krump and You 2018; Martin et al. 2018) that infects an estimated five to ten million people worldwide (Gessain and Cassar 2012). Viral integration into host T lymphocytes, typically mature CD4[+] T-cells, results in adult T-cell leukemia/lymphoma and HTLV-1-associated myelopathy/tropical spastic paraparesis in ~5% and ~1% of individuals, respectively (Poiesz et al. 1980; Yoshida et al. 1982; Gessain et al. 1985; Osame et al. 1986; Matsuoka and Jeang 2007; Bangham et al. 2015; Futsch

et al. 2017; Bangham 2018). New HTLV-1 infections are established when the viral positive-sense, single-stranded RNA genome is reverse transcribed and the resulting double-stranded DNA is integrated into the host genome (Fan et al. 1992; Jones et al. 2008; Carpentier et al. 2015; Martin et al. 2016; Meissner et al. 2017). Both processes require viral enzymes that are expressed by way of two independent −1 PRFs (Fig. 1A; Mador et al. 1989; Nam et al. 1993). Therefore, −1 PRFs are essential to HTLV-1 infection. Although −1 PRF has been extensively studied in a number of other viruses (for review, see Farabaugh 1996; Gesteland and Atkins 1996; Brierley and Dos Ramos 2006; Giedroc and Cornish 2009; Liao et al. 2011; Dinman 2012b; Caliskan et al. 2015; Yan et al. 2015; Atkins et al. 2016), it is less well studied in HTLV-1.

HTLV-1 contains two −1 PRF sites in its viral genome (Fig. 1A). These sites (*gag-pro* and *pro-pol*) are required for the translation of its *pro* and *pol* genes. In this work, we focus on the HTLV-1 *pro-pol* frameshift site (Fig. 1B). Although the slippery sequence (UUUAAAC) is known (Nam et al. 1993) and a pseudoknot structural model was proposed (Supplemental Fig. S1; Brierley et al. 1989; ten Dam et al. 1990; Le et al. 1991), no experimental data substantiating this structure is published, nor has the
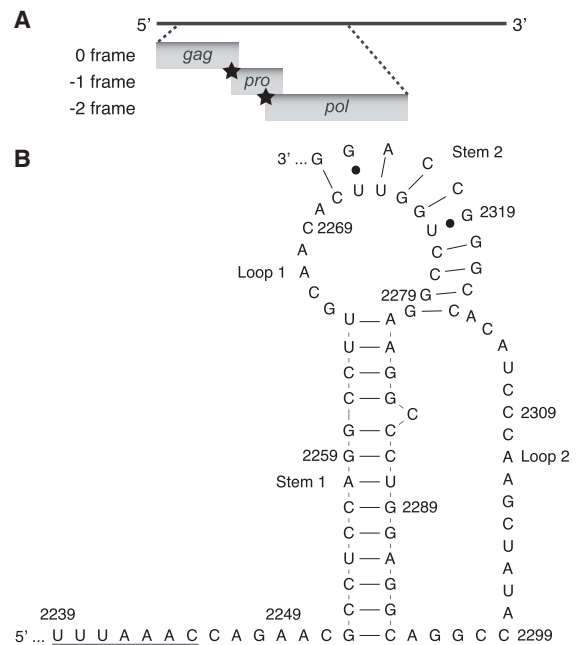

FIGURE 1. HTLV-1 programmed ribosomal frameshift sites. (*A*) A simplified cartoon of the HTLV-1 RNA genome is shown. −1 PRF sites (starred) are used for the translation of the *pro* and *pol* open reading frames, which are in the −1 and −2 reading frames relative to *gag*, respectively. (*B*) The HTLV-1 *pro-pol* frameshift site includes a slippery sequence (underlined), a spacer (nucleotides 2246–2251), and a downstream structure, which was proposed to be an H-type pseudoknot. The nucleotide numbering is derived from the NCBI reference sequence NC_001436.1.

frameshift efficiency been precisely measured. The lack of structural data is particularly problematic, as it has led to confusion over what defines the frameshift site's 3′ boundary. For example, the most recent investigation of HTLV-1 *pro-pol* −1 PRF mechanism used a frameshift site that only included nucleotides 2239–2299 (NCBI reference sequence NC_001436.1) (Liao et al. 2011). This 3′ boundary limits the frameshift site structure to a stem–loop. However, if the frameshift site includes one of the previously proposed pseudoknot structures (Supplemental Fig. S1; Brierley et al. 1989; ten Dam et al. 1990; Le et al. 1991), then the 3′ end of the frameshift site should be defined by nucleotide 2322 or 2324 (Fig. 1B).

The objective of our work was to conclusively define the RNA elements within and the 3′ boundary of the HTLV-1 *pro-pol* frameshift site. Towards these objectives, we began by quantifying the in vitro frameshift efficiency of three HTLV-1 *pro-pol* frameshift site constructs that had different 3′ boundaries. Our results indicated that nucleotides included in the longest construct were essential to highly efficient frameshift stimulation. Interestingly, only this construct could form the putative frameshift site pseudoknot. Next, we determined the secondary structure of this frameshift site. We found that the dominant structure was an H-type pseudoknot, the sequence of which is highly conserved. The pseudoknot's critical role in frameshift stimulation was directly revealed by examining the impact of structural changes on HTLV-1 *pro-pol* frameshift efficiencies. As predicted, mutations that occluded pseudoknot formation drastically reduced the frameshift efficiency. These results are significant because they demonstrate that an H-type pseudoknot is important to HTLV-1 *pro-pol* −1 PRF and define the frameshift site 3′ boundary

as nucleotide 2324 (Fig. 1B). This work expands our understanding of HTLV-1 *pro-pol* −1 PRF and establishes a firm foundation for its future investigation.

RESULTS

Measurement of the HTLV-1 *pro-pol* frameshift efficiency

In the most recent investigation of the HTLV-1 *pro-pol* −1 PRF mechanism, the frameshift site 3′ boundary was set at nucleotide 2299 (Liao et al. 2011). This boundary restricts the frameshift site structure to a stem–loop. However, if the frameshift site includes one of the previously proposed pseudoknot structures (Supplemental Fig. S1; Brierley et al. 1989; ten Dam et al. 1990; Le et al. 1991), then the 3′ boundary should be described by nucleotide 2322 or 2324 (Fig. 1B). The frameshift efficiency for a HTLV-1 *pro-pol* frameshift site that includes nucleotides 2299–2324 has not been reported. Therefore, the importance of nucleotides 2299–2324 to frameshifting are unknown. We hypothesized that the downstream RNA was critical to frameshifting because it can base-pair with loop nucleotides to form the putative pseudoknot structure.

To test our hypothesis, we quantified the HTLV-1 *pro-pol* frameshift efficiency for three frameshift sites that differed in the length of their 3′ ends. The level of frameshift stimulation produced by each variant was measured using a well-established dual-luciferase frameshift assay (Fig. 2A; Grentzmann et al. 1998). The HTLV-1 *pro-pol* constructs were designed such that the longest construct (wild-type, WT) could form the putative pseudoknot, but the shorter constructs (pseudoknot deletion mutants 1 and 2,
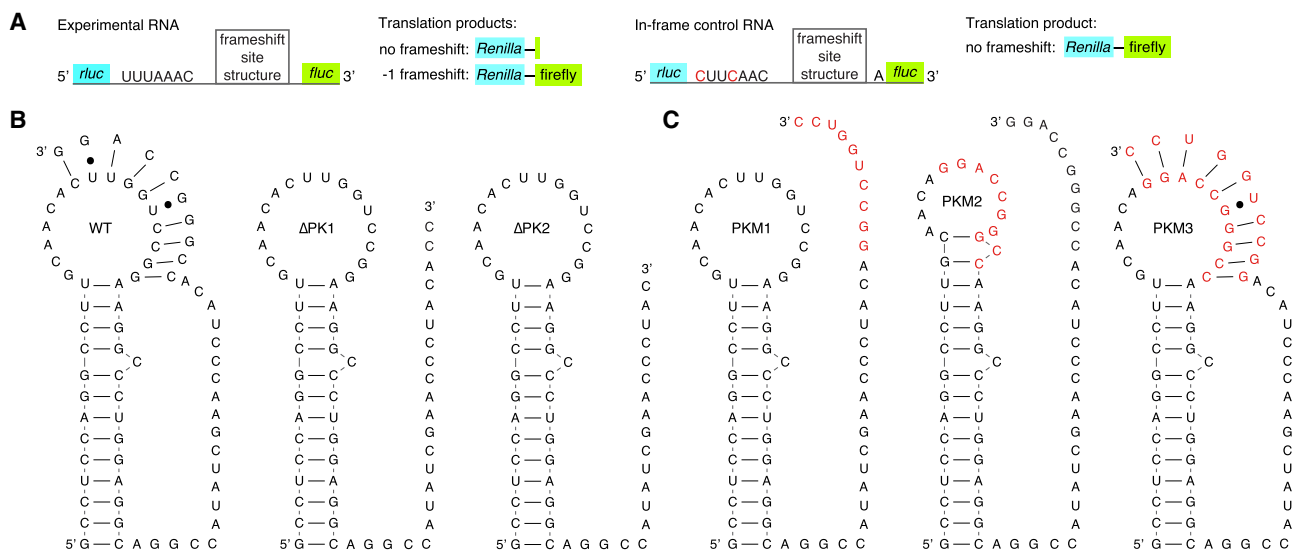

FIGURE 2. Dual-luciferase frameshift assay constructs. (*A*) A schematic of the RNAs used to quantify the HTLV-1 *pro-pol* −1 PRF efficiency and their associated translation products. Expected frameshift site structures of the (*B*) WT, ΔPK1, and ΔPK2 constructs, and (*C*) of the PKM1, PKM2, and PKM3 constructs.

ΔPK1 and ΔPK2) could not (Fig. 2B). We report a HTLV-1 *pro-pol* frameshift efficiency of 19.4(±0.3)% for the WT frameshift site (Table 1). In contrast, the ΔPK1 and ΔPK2 frameshift efficiencies were 2.98(±0.06)% and 4.93(± 0.04)%, respectively. Unpaired two-sample *t*-tests were used to determine if the frameshift efficiencies measured were statistically different from the WT frameshift efficiency. The *P*-values demonstrate that the HTLV-1 ΔPK1 and ΔPK2 in vitro frameshift efficiencies were statistically different from the WT frameshift efficiency (Table 1). The in vitro frameshift efficiencies of ΔPK1 and ΔPK2 are comparable in magnitude to what was previously reported for a similarly truncated HTLV-1 *pro-pol* frameshift site (~5%) (Liao et al. 2011). The relative increase in frameshift efficiency provided by the WT construct suggests that the additional nucleotides included in its 3′ end are essential to −1 PRF.

The in vitro frameshift efficiencies for three control frameshift sites were also measured (Table 1). The well-characterized HIV-1 −1 PRF site was used as a positive control. Two negative controls, ΔPK3 and an out-of-frame background control, were designed for this study. The ΔPK3 frameshift site included the WT HTLV-1 *pro-pol* slippery sequence and spacer, but lacked a frameshift site structure (nucleotides 2252–2326 were deleted). A −1 PRF in this context would be induced by the slippery sequence (UUUAAAC) without the aid of a structure, as the sequence coding for the pseudoknot (2252–2324) is removed. The out-of-frame background control was identical to ΔPK3, except that it also included slippery sequence mutations (UUUAAAC was changed to CUUCAAC) that made the frameshift site nonfunctional. Given that this control lacks a functional frameshift site, a −1 PRF in this context demonstrated the

background rate of nonprogrammed ribosomal frameshifting. Overall, the in vitro frameshift efficiencies for these three controls were consistent with previous publications. The HIV-1 positive control in vitro frameshift efficiency (6.84[±0.05]%) was similar to previous in vitro measurements (5.6[±0.4]% [Kim et al. 2001], ~4.7% [Marcheschi et al. 2009], 4.6[±0.5]% [Mouzakis et al. 2013], 5.4[±0.9]% [Low et al. 2014], and ~7% [Meydan et al. 2017]). The ΔPK3 negative control in vitro frameshift efficiency (0.850[±0.007]%) was comparable to the frameshift efficiency reported for a UUUAAAC slippery sequence alone (0.6[±0.1]%) (Yu et al. 2010). Lastly, the frequency of nonprogrammed ribosomal frameshifting (0.209[±0.002]%), which was measured with the out-of-frame background control, was close to the ~0.3% in vitro background frameshift efficiency reported for the p2luc system (Grentzmann et al. 1998).

The HTLV-1 *pro-pol* WT frameshift site secondary structure was determined by chemical probing and computational modeling

We hypothesized that the nucleotides at the 3′ end of the WT frameshift site were critical to frameshift stimulation because they can base-pair with RNA upstream to form a pseudoknot structure (Fig. 1B). To test this hypothesis, the frameshift site secondary structure was determined. While secondary structure prediction for simple RNA structures is fairly accurate, predictions of complex RNA structures, such as pseudoknots, have limited accuracy in the absence of experimental data (Hajdin et al. 2013). Therefore, we sought to combine experimental data with computational tools to determine the HTLV-1 *pro-pol*

TABLE 1. In vitro frameshift efficiencies

	Frameshift efficiency ± standard error (%)	Relative change in frameshift efficiency[a]	Statistically significant? (*P*-value[b])
HTLV-1 *pro-pol* frameshift site variant			
WT	19.4 ± 0.3	n/a	
ΔPK1	2.98 ± 0.06	−85%	Yes (6.79 × 10⁻¹¹)
ΔPK2	4.93 ± 0.04	−75%	Yes (1.54 × 10⁻¹⁰)
PKM1	2.52 ± 0.03	−87%	Yes (5.08 × 10⁻¹¹)
PKM2	9.5 ± 0.1	−51%	Yes (5.06 × 10⁻¹¹)
PKM3	28.7 ± 0.3	+48%	Yes (1.54 × 10⁻¹⁰)
Positive control			
HIV-1	6.84 ± 0.05	n/a	Yes (4.29 × 10⁻¹⁰)
Negative controls			
ΔPK3	0.850 ± 0.007	−96%	Yes (2.47 × 10⁻¹¹)
Out-of-frame background control	0.209 ± 0.002	−99%	Yes (1.09 × 10⁻¹²)

[a]Calculated by dividing the variant's frameshift efficiency by WT's. Negative and positive values reflect decreases and increases in frameshift efficiency, respectively.
[b]*P*-values calculated using unpaired two-sample *t*-tests with 95% confidence levels. Not applicable (n/a).

frameshift site secondary structure. Extremely accurate secondary structures of large and complex RNAs, which include pseudoknots, were determined using this approach (Merino et al. 2005; Wilkinson et al. 2006; Reuter and Mathews 2010; Hajdin et al. 2013; Leonard et al. 2013; Chapman et al. 2014; Cordero et al. 2014a; Tian et al. 2014; Hartwick et al. 2018).

The HTLV-1 *pro-pol* frameshift site (nucleotides 2239–2324) was placed into an RNA cassette compatible with chemical modification and primer extension experiments (Supplemental Fig. 2A; Cordero et al. 2014a; Tian et al. 2014). In vitro transcribed and denaturing polyacrylamide gel electrophoresis (PAGE) purified RNA was refolded in a magnesium-containing buffer. Folding conditions were optimized to ensure that the RNA was monomeric (Materials and Methods). Dimer formation was not observed when the RNA was refolded. Next, the RNA was chemically probed with molecules that selectively modify single-stranded RNA. Specifically, N-methylisatoic anhydride (NMIA) covalently modifies the ribose 2′ OH of single-stranded, unconstrained nucleotides (Merino et al. 2005; Weeks and Mauger 2011). Similarly, dimethyl sulfate (DMS) covalently modifies the N1/N3 of A/C bases of single-stranded nucleotides, respectively (Peattie and Gilbert 1980; Ehresmann et al. 1987; Tijerina et al. 2007; Weeks 2010). The chemically modified RNA was reverse transcribed and the complementary DNA library was analyzed with capillary electrophoresis (CE) (Supplemental Fig. S2D). Nucleotide reactivities were normalized, quantified, and averaged (Supplemental Fig. S2B,C), as previously described (Kladwang et al. 2011a, 2014; Yoon et al. 2011; Kim et al. 2013; Tian et al. 2014; Lee et al. 2015; Hartwick et al. 2018).

The NMIA nucleotide reactivities were used as a constraint within the RNAstructure Shapeknots secondary structure prediction algorithm (Reuter and Mathews 2010) to generate a data-driven secondary structure of the HTLV-1 *pro-pol* frameshift site (Fig. 3A). The dominant secondary structure derived was an H-type pseudoknot. This structure had extremely high (97%–100%) helix-wide confidence values (Fig. 3A) and is well supported by the corresponding DMS chemical probing results (Fig. 3B).

H-type, or hairpin loop type, pseudoknots are RNA structures that form when the loop nucleotides within a stem–loop base pair with complementary nucleotides downstream (Peselis and Serganov 2014). Similar to the majority of naturally occurring H-type pseudoknots, the HTLV-1 *pro-pol* frameshift site pseudoknot (Fig. 3) has two helices, which could coaxially stack, separated by two loops. Indeed, the HTLV-1 *pro-pol* frameshift site structure was previously proposed to be an H-type pseudoknot by three independent groups (Brierley et al. 1989; ten Dam et al. 1990; Le et al. 1991). In these models, the lengths of stem 1 (13 bp) and loop 2 (20 nt) were consistent, but the lengths of loop 1 (6–8 nt) and stem 2 (8 or 10 bp) varied (Supplemental Fig. S1). The HTLV-1 *pro-pol* pseudoknot described here has 13 bp in stem 1, 6 nt in loop 1, 10 bp in stem 2, and 20 nt in loop 2 (Fig. 3). This structure is consistent with one (Supplemental Fig. S1B; ten Dam et al. 1990) of the three previously described frameshift site structural models.

The HTLV-1 *pro-pol* −1 PRF site is highly conserved

The importance of pseudoknot formation for HTLV-1 replication may be reflected by a high level of sequence conservation. The HTLV-1 reference sequence used in this
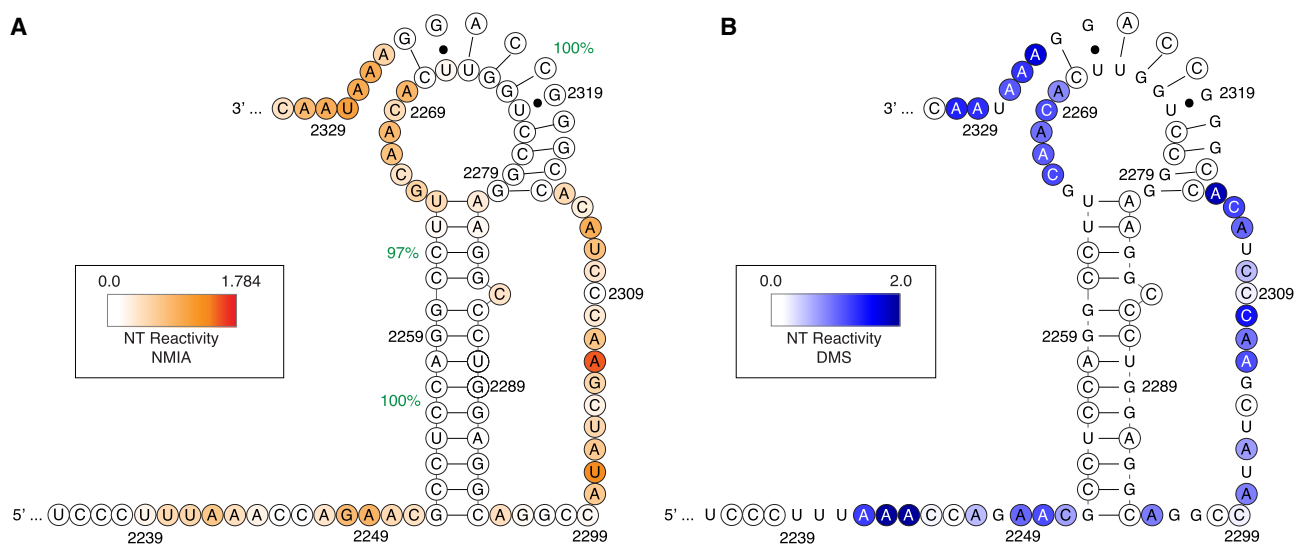

FIGURE 3. The HTLV-1 frameshift site structure is a pseudoknot. (A) The data-driven secondary structure of the HTLV-1 *pro-pol* frameshift site is shown. Percentages represent helix-wide confidence values determined with bootstrapping methods. Normalized and averaged NMIA nucleotide (NT) reactivities are overlaid onto the structure. Numbering is consistent with the NCBI reference sequence NC_001436.1 and Figure 1B. (B) Normalized and averaged A and C NT reactivities (derived from experiments with DMS) are overlaid onto the structure in A.
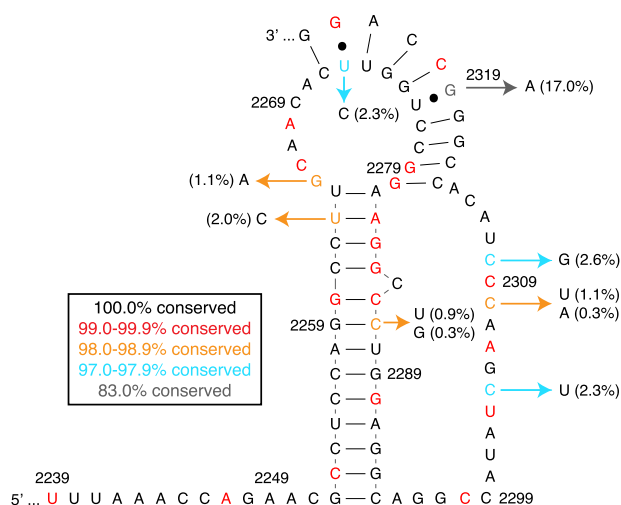
FIGURE 4. The HTLV-1 *pro-pol* frameshift site sequence is highly conserved. The consensus sequence derived from 352 HTLV-1 sequences available through the NCBI (as of January 13, 2019) is highly conserved and identical to the NCBI reference sequence NC_001436.1. Nucleotides are shown in different colors according to their sequence conservation. Nucleotides 100% conserved are black, 99.0%–99.9% conserved are red, 98.0%–98.9% conserved are orange, 97.0%–97.9% are cyan, and the single nucleotide 83.0% conserved is gray. The second most common nucleotide is shown for nucleotides that are <99.0% conserved. Nucleotide numbering is consistent with Figure 1B.

work, NC_001436.1, has a *pro-pol* frameshift site sequence identical to the consensus sequence derived from 352 HTLV-1 sequences available through the NCBI. Alignment of those sequences revealed that the frameshift site and its pseudoknot are both highly conserved (Fig. 4). Although no significant base-pair covariation was detected using this alignment, most of the variation does not appear to affect the frameshift site structure (Fig. 4). A lack of base-pair covariation is not surprising given the high-level of pseudoknot sequence conservation (66 of 73 nt are ≥99% conserved). Limited sequence variance is also expected because the pseudoknot codes for proteins in alternate reading frames, is part of a −1 PRF site, and is derived from a retrovirus with relatively low genetic drift (Gessain et al. 1992; Van Dooren et al. 2004).

Mutations to the HTLV-1 *pro-pol* frameshift site demonstrate that a pseudoknot is important for highly efficient frameshift stimulation

Having determined that the frameshift site structure is a highly conserved H-type pseudoknot and inclusion of its whole sequence was essential to −1 PRF, we sought to directly demonstrate that the pseudoknot was responsible for the 19.4(±0.3)% frameshift efficiency observed. We hypothesized that sequence changes that eliminate pseudoknot formation while preserving the frameshift site's

sequence length should decrease the −1 frameshift efficiency. To test this hypothesis, we measured the frameshift efficiency of three variant frameshift sites designed to modify the pseudoknot structure without changing its 3′ boundary (Fig. 2C). In these pseudoknot mutants (PKM1, PKM2, and PKM3), sequence changes were made to alter base-pairing within stem 2 (Fig. 1B). For the PKM1 and PKM2 frameshift sites, nucleotides 2315–2324 (PKM1) and 2271–2280 (PKM2) were modified to prevent stem 2 base-pairing and preserve their sequence lengths (Fig. 2C). In contrast, the PKM3 frameshift site combines the PKM1 and PKM2 mutations to restore the pseudoknot through compensatory mutations.

Frameshift efficiencies for PKM1-3 were measured using the dual-luciferase assay described earlier. Consistent with our hypothesis, the in vitro frameshift efficiencies of the PKM1 (2.52[±0.03]%) and PKM2 (9.5[±0.1]%) frameshift sites were strikingly reduced relative to the WT frameshift site (19.4[±0.3]%) (Table 1). PKM1 and PKM2 were both expected to form stem–loop structures of slightly different lengths (Fig. 2C). While these structures, which were ultimately determined as reported in the next section, do stimulate −1 PRF to a certain extent, they are less effective than the WT pseudoknot. This indicates that the pseudoknot plays an important role in inducing −1 PRF. In support of this idea, PKM3 increased the frameshift efficiency (28.7[±0.3]%) relative to the WT level and was designed to form a pseudoknot through compensatory mutations (Fig. 2C).

The PKM1 frameshift efficiency (2.52[±0.03]%) was similar to ΔPK1's (2.98[±0.06]%) and ΔPK2's (4.93[±0.04]%). Given that the ΔPK1 and ΔPK2 frameshift sites lack the 3′ nucleotides required to form a pseudoknot, the PKM1 mutations appeared to eliminate pseudoknot formation as designed. Even though PKM2 was also expected to form a stem–loop structure (Fig. 2C), its frameshift efficiency (9.5[±0.1]%) was surprisingly high relative to ΔPK1, ΔPK2, and PKM1 (Table 1). We hypothesized that PKM2's 2 bp extension might explain the observed differences in frameshift stimulation. The PKM3 variant increased frameshifting by ∼1.5-fold relative to the WT frameshift site (28.7[±0.3]% vs. 19.4[±0.3]%). This difference may be explained by the increased thermodynamic stability provided by a G-U to G-C base-pair substitution in stem 2 (Fig. 2C), as well as by possible differences in tertiary structure.

The secondary structures of the frameshift site mutants are consistent with their impact on frameshift stimulation

To confirm that the ΔPK1, PKM1, PKM2, and PKM3 pseudoknot mutants from our functional assays altered the pseudoknot structure as intended (Fig. 2B,C), we determined their secondary structures. By following the chemical modification and computational approach described above, dominant secondary structures were derived for

each mutant (Fig. 5; Supplemental Figs. S3–S6). All three mutants designed to disrupt the pseudoknot (ΔPK1, PKM1, and PKM2) formed stem–loop structures instead of pseudoknots (Fig. 5A–C). As anticipated, the PKM3 sequence restored the pseudoknot structure through compensatory mutations (Fig. 5D).

The structures of the ΔPK1, PKM1, and PKM2 variants also help clarify their unexpected differences in −1 PRF. The ΔPK1 and PKM1 variants were expected to have the same frameshift site structure (Fig. 2B,C) and thus, frameshift efficiency. A close examination of their secondary structures (Fig. 5A,B) reveals the stem–loop structures downstream from the slippery sequence were indistinguishable. While PKM1 formed a second stem–loop 3′ of the first (Fig. 5B), this is not expected to significantly influ-

ence −1 PRF. This expectation is supported by the ΔPK1 variant, which does not have a second downstream stem–loop and stimulates frameshifting to a similar extent as PKM1 (Table 1). The base-paired region of the PKM2 stem–loop was extended by 2 bp (Fig. 5C), as expected (Fig. 2C). Formation of two extra base pairs could change the folding kinetics of the stem–loop and therefore potentially increase the frameshift efficiency (Yang et al. 2018). Thus, this difference in PKM2's structure (Fig. 5C) may explain the surprising difference in frameshift stimulation observed for the PKM2 variant in comparison to the ΔPK1 and PKM1 variants (Table 1).

The structures of PKM3 and WT were nearly identical (Fig. 3 vs. Fig. 5D). Yet, PKM3's frameshift efficiency was ~1.5-fold higher than WT's (Table 1). We suspect that
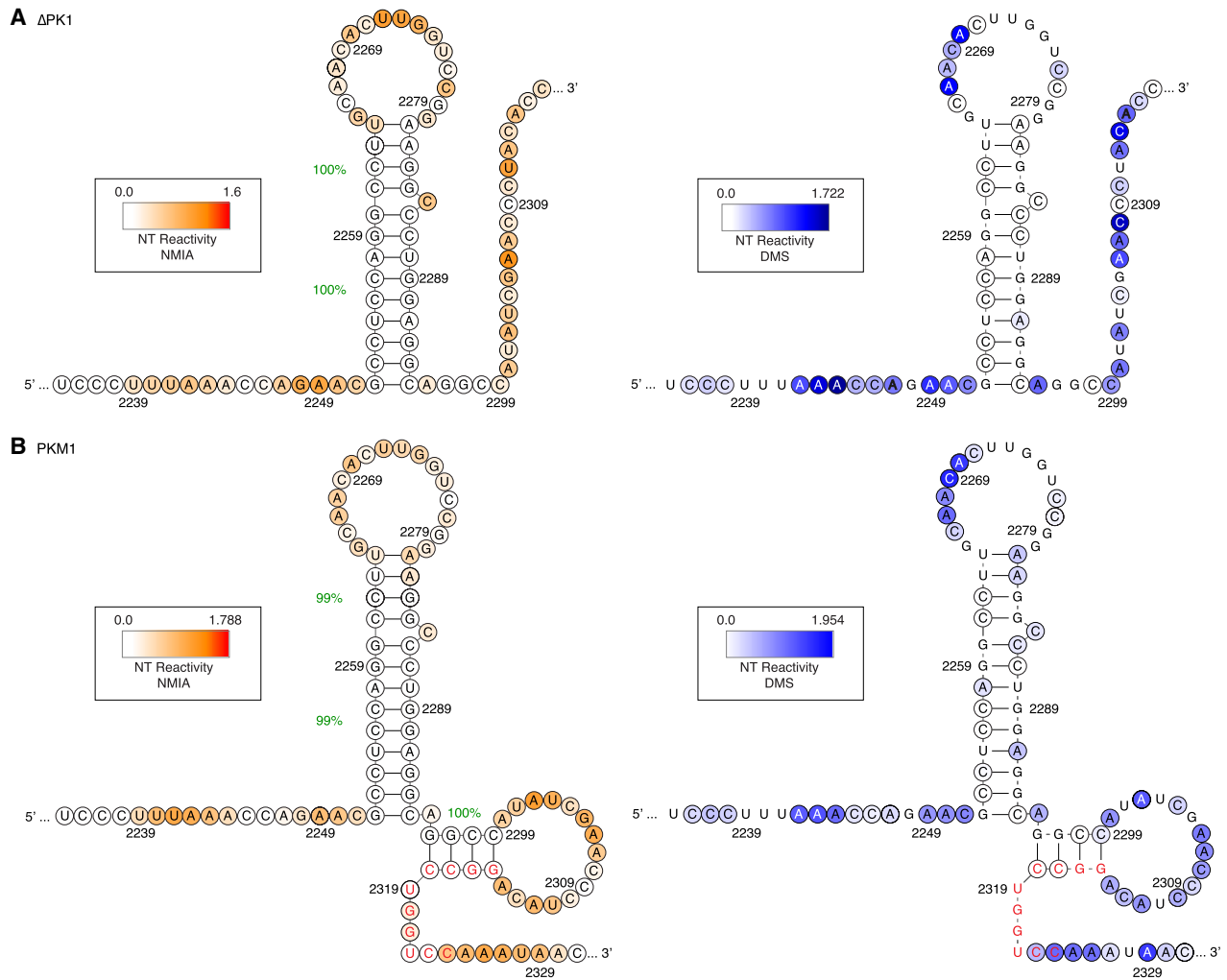

FIGURE 5. Secondary structures of the HTLV-1 *pro-pol* frameshift site variants. The data-driven secondary structures of the HTLV-1 *pro-pol* (*A*) ΔPK1, (*B*) PKM1, (*C*) PKM2, and (*D*) PKM3 frameshift site variants are shown. On the *left* side of the panel, the normalized and averaged NMIA nucleotide (NT) reactivities are superimposed on the structures. Percentages shown represent helix-wide confidence values determined with bootstrapping methods. On the *right* side of the panel, normalized and averaged A and C NT reactivities (derived from experiments with DMS) are overlaid onto each structure. All sequence changes from WT are shown in red font. Note that 2 bp predicted in the loops of ΔPK1 and PKM1 between nucleotides A2268–U2276 and C2269–G2275 were not included in the figures. (*Figure continues on next page.*)
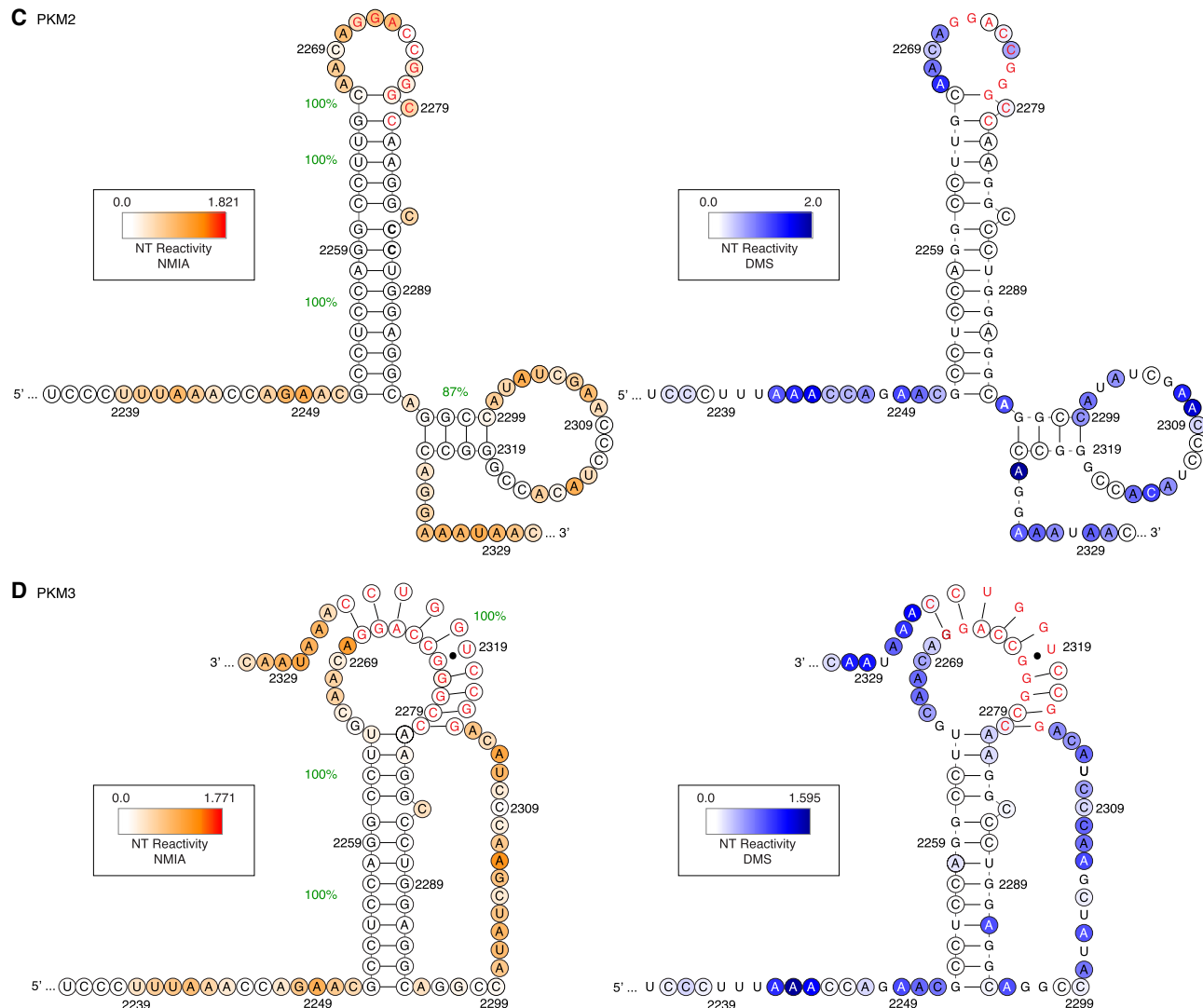
FIGURE 5. (*Continued*)

this increase might result from differences in pseudoknot thermodynamic stability, provided by a G-U to G-C base-pair substitution in stem 2 or a possible extension of stem 1 akin to what was observed for PKM2. An extension of the PKM3 stem 1 by 1 bp (formed between G2265 and C2280) would remain consistent with the chemical probing data (Fig. 5D) even if it disrupted the C2280 and G2315 base pair in stem 2. Given that an extension of the PKM2 stem 1 by 2 bp increased frameshifting relative to ΔPK1 and PKM1, perhaps a similar situation contributes to the ~9% difference in frameshift efficiency between the PKM3 and WT frameshift sites.

DISCUSSION

In this work, we aimed to conclusively define the HTLV-1 *pro-pol* frameshift site's RNA structure and 3′ boundary. We hypothesized that the frameshift site structure was a

pseudoknot and that the frameshift site 3′ boundary would be defined by the 3′ end of that pseudoknot (Fig. 1B). A well-established dual-luciferase frameshift assay (Grentzmann et al. 1998) was used to measure the in vitro frameshift efficiency of several HTLV-1 *pro-pol* variants. This assay was reliably used to measure the frameshift efficiency of the related HIV-1 retrovirus (Marcheschi et al. 2009; Mouzakis et al. 2013; Low et al. 2014), which also replicates within T-cells. Three HTLV-1 *pro-pol* frameshift site variants (WT, ΔPK1, and ΔPK2) were designed to examine how differences in 3′ boundary affected frameshift efficiency. Three additional HTLV-1 *pro-pol* variants (PKM1-3) were designed to evaluate the importance of pseudoknot formation to frameshift stimulation. The WT, ΔPK1, and PKM1-3 secondary structures were determined by combining data from chemical probing experiments with computational folding algorithms (Reuter and Mathews 2010). Our results indicate that the HTLV-1 *pro-*

pol frameshift site structure is that of a classic H-type pseudoknot (Fig. 3). This pseudoknot acts in *cis* with the slippery sequence to stimulate −1 PRF to 19.4(±0.3)%. To the best of our knowledge, this is the first quantitative measurement of the HTLV-1 *pro-pol* −1 frameshift efficiency for a frameshift site that includes the pseudoknot structure. Frameshift stimulation by both the WT and PKM3 pseudoknots was much greater than that produced by the stem–loop structures of ΔPK1, PKM1, and PKM2 (Table 1). These results are significant because, for the first time, they demonstrate that a pseudoknot is important to HTLV-1 *pro-pol* −1 PRF and define the 3′ boundary of the HTLV-1 *pro-pol* frameshift site as nucleotide 2324.

There is substantial sequence conservation (79%, calculated using NCBI BLASTn) between the HTLV-1 (NC_001436.1, nucleotides 2239–2324) and the HTLV type 2 (HTLV-2, NC_001488.1 nucleotides 2587–2672) *pro-pol* frameshift sites. Therefore, we would expect the HTLV-1 frameshift efficiency to be similar in magnitude to that measured for this site in HTLV-2. Research by Mador et al. demonstrated that when the HTLV-2 *pro-pol* frameshift efficiency is measured in absence of the upstream *gag-pro* frameshift site, the HTLV-2 *pro-pol* frameshift efficiency is approximately 27% (Mador et al. 1989). This value is comparable to the HTLV-1 *pro-pol* frameshift efficiency reported here (19.4[±0.3]%), which was also measured in the absence of an upstream *gag-pro* frameshift site. Variations in approach likely contributed most to the difference in frameshift efficiency observed. While both studies used an in vitro rabbit reticulocyte translation system, the RNA constructs translated were fairly distinct. The HTLV-2 frameshift site was surrounded by HTLV-2 RNA (Mador et al. 1989), while the HTLV-1 frameshift site was flanked by RNA coding for luciferase enzymes. Differences in sequence context that impact local secondary structure can change the −1 frameshift efficiency, as was demonstrated for the HIV-1 frameshift site (Low et al. 2014). Disparities in relative mRNA to ribosome concentrations or translation initiation rates would also alter the ribosome density on the transcripts. Previous studies have also shown that higher translation initiation rates, which presumably increase ribosome density, cause a reduction in −1 PRF (Honigman et al. 1995; Lopinski et al. 2000; Gendron et al. 2008; Charbonneau et al. 2012). Given the variance in frameshift site sequence, the difference in sequence context surrounding the frameshift sites, and the minor, but probable differences in experimental design, it is not surprising that a moderate difference in the HTLV-2 and the HTLV-1 *pro-pol* frameshift efficiency was observed.

Interestingly, when the HTLV-2 *pro-pol* frameshift site was translated in the presence of the upstream frameshift site, three polyproteins are produced (Gag, Gag-Pro, and Gag-Pro-Pol) at a relative frequency of 100:13:0.9 (Mador et al. 1989). This ratio corresponds to a HTLV-2

pro-pol frameshift efficiency of ~7%, which was reduced relative to what was measured in the absence of a functional *gag-pro* frameshift site (~27%) in the same study (Mador et al. 1989). While beyond the scope of this study, it will be interesting to determine whether the HTLV-1 *pro-pol* frameshift site stimulates frameshifting to the same degree when the upstream *gag-pro* frameshift site is included.

When evaluating the importance of the HTLV-1 *pro-pol* pseudoknot structure to frameshift site function, we serendipitously designed a frameshift site variant (PKM3) with a substantially higher frameshift efficiency (28.7[±0.3]% vs. 19.4[±0.3]%). This could imply that the HTLV-1 *pro-pol* frameshift site evolved in a way that limited the frameshift efficiency to what was necessary for HTLV-1 replication. In the context of HTLV-1, the *pro-pol* frameshift efficiency regulates the molar ratio of the viral protease enzyme to the reverse transcriptase and integrase enzymes. Access to the *pro* open reading frame, which codes for protease, is further regulated by the upstream *gag-pro* frameshift site and its associated frameshift efficiency (Fig. 1A; Mador et al. 1989; Nam et al. 1993). Control of HTLV-1 protein ratios through −1 PRF is consistent with the "golden-mean" model previously described (Plant et al. 2010). In this model, viruses use mechanisms of translational reprogramming, such as PRF and translational attenuation, to modulate the relative ratios of their encoded proteins (Plant et al. 2010; Tholstrup et al. 2012; Huang et al. 2013). Our results may provide yet another example of a "tuned" −1 PRF efficiency.

As this is the first conclusive report of the HTLV-1 *pro-pol* frameshift site structure and associated frameshift efficiency, a comparison of its structure and function to other well-characterized −1 PRF sites is warranted. Several viruses, including retroviruses, coronaviruses, and luteoviruses, also use classical H-type pseudoknot structures to stimulate −1 PRF (Giedroc et al. 2000; Staple and Butcher 2005; Brierley et al. 2008; Giedroc and Cornish 2009). In Table 2, the frameshift efficiencies and pseudoknot stem and loop lengths are listed for several well-characterized viruses of this type. Two salient features are revealed by this juxtaposition. First, the HTLV-1 *pro-pol* frameshift efficiency is similar in magnitude to the frameshift efficiencies of the other viruses. Second, the HTLV-1 pseudoknot appears to be substantially larger than its counterparts. If the HTLV-1 pseudoknot stem 1 and stem 2 were coaxially stacked, the stacked helix would include 23 bp. In contrast, the Infectious Bronchitis Virus (IBV) pseudoknot, which is the largest in the comparator set (Table 2), would only include 17 bp. The HTLV-1 loop 2 length (20 nt) is also much longer than the rest, which vary in length (7–14 nt). Intriguingly, changes to the IBV pseudoknot stem 1 length (Napthine et al. 1999) and the Simian Retrovirus Type 1 (SRV-1) loop 2 length (ten Dam et al. 1995) significantly impacted their −1 frameshift efficiencies. The importance of

TABLE 2. A comparison of the HTLV-1 *pro-pol* frameshift efficiency and structure to well-characterized −1 PRF sites that include H-type pseudoknots

Virus type	Frameshift site	Frameshift efficiency	Pseudoknot stem and loop lengths				
			S1	L1	ISE	S2	L2
Retrovirus	HTLV-1 *pro-pol*	19.4(±0.3)%	13 bp	6 nt	0 nt	10 bp	20 nt
Retrovirus	MMTV *gag-pro*	~20% (Chamorro et al. 1992)	5 bp	1 nt	1 nt	7 bp	8 nt
Retrovirus	SRV-1 *gag-pro*	23% (Michiels et al. 2001)	6 bp	1 nt	1 nt	6 bp	12 nt
Retrovirus	VMV *gag-pol*	28.3% (Pennell et al. 2008)	7 bp	5 nt	7 nt	7 bp	14 nt
Coronavirus	IBV *pp1a-1b*	~30% (Kontos et al. 2001)	11 bp	2 nt	0 nt	6 bp	8 nt
Luteoviruses	BWYV *P1-P2*	6(±1)% (Cornish et al. 2005)	5 bp	2 nt	1 nt	3 bp	7 nt
Luteoviruses	ScYLV *P1-P2*	15(±2)% (Cornish et al. 2005)	5 bp	2 nt	1 nt	3 bp	9 nt

Viruses: HTLV-1, Human T-cell Leukemia Virus Type 1; IBV, Infectious Bronchitis Virus; BWYV, Beet Western Yellow Virus; MMTV, Mouse Mammary Tumor Virus; ScYLV, Sugarcane Leaf Yellow Virus; VMV, Visna-Maedi virus; SRV-1, Simian Retrovirus Type 1; Pseudoknot descriptors: S1, stem 1; L1, loop 1; ISE, interstem element; S2, stem 2; L2, loop 2.

the HTLV-1 pseudoknot stem and loop lengths to −1 PRF have not been examined. It will be interesting to see how these types of changes impact the HTLV-1 *pro-pol* frameshift efficiency.

Base-triples are commonly observed in H-type pseudoknot structures and form when loop nucleotides hydrogen bond with the non-Watson–Crick faces of the bases in the minor (stem 1/loop 2) and major (stem 2/loop 1) grooves (Giedroc et al. 2000; Staple and Butcher 2005; Giedroc and Cornish 2009; Butcher and Pyle 2011; Peselis and Serganov 2014; Devi et al. 2015). These tertiary interactions impact pseudoknot mechanical stability (Chen et al. 2009, 2017), helical junction geometry (Gupta and Bansal 2014), and conformational plasticity (Ritchie et al. 2012; Wu et al. 2018; Halma et al. 2019). As a consequence, these tertiary contacts can be critical to frameshift stimulation. For example, the SRV-1 pseudoknot has an extensive triplex formed between its adenosine-rich loop 2 and G-C-rich stem 1 minor groove (Michiels et al. 2001). When the loop 2 adenosines are replaced with pyrimidines, which disrupt the triplex, the SRV-1 frameshift efficiency is reduced by 88% (Olsthoorn et al. 2010). Base-triples are typically identified with atomic techniques, such as NMR and X-ray crystallography (Giedroc et al. 2000; Staple and Butcher 2005; Cornish et al. 2006; Brierley et al. 2008; Giedroc and Cornish 2009; Peselis and Serganov 2014; Chen et al. 2017). While not definitive, one-dimensional chemical probing experiments could also identify loop nucleotides forming base-triples in a pseudoknot. Theoretically, if a loop nucleotide is structurally constrained as a result of a tertiary interaction, its 2′OH might have limited flexibility and its base might have decreased solvent accessibility. A constrained 2′OH would limit that nucleotide's reactivity with NMIA and a decreased base solvent accessibility would reduce that nucleotide's reactivity with DMS (Weeks 2010). Therefore, loop nucleotides that do not react with NMIA and DMS

might be part of a base-triple. One caveat to this analysis is that it depends on low background signal from the reverse transcription no-modification controls. While our background signal was generally low, strong reverse transcriptase stops were observed in several of our data sets, including those for the WT (Supplemental Fig. S2D) and PKM3 (Supplemental Fig. S7D) pseudoknots. Therefore, our chemical probing data is not well suited for the identification of HTLV-1 *pro-pol* pseudoknot base-triples. Experiments outside the scope of this work are necessary to precisely describe these and other unidentified tertiary interactions. Ultimately, an atomic-resolution structure is needed to begin the process of distinguishing which of the HTLV-1 *pro-pol* pseudoknot's features are critical to structure-stimulated frameshifting.

The HTLV-1 *pro-pol* RNA secondary structures reported here (Figs. 3, 5) were determined using RNAs that were in vitro transcribed, purified under denaturing conditions, and subsequently refolded. While this methodology is frequently used for the synthesis and purification of large quantities of RNA used in structural research, there are limitations that arise from this approach. One such limitation is that the structure of the refolded RNA may be different from what would be found *in vivo*. Whether the HTLV-1 *pro-pol* frameshift site structure determined here is consistent with the structure forming inside the cytosol of CD4[+] T-cells, which is where the ribosome would encounter it, is yet to be determined.

MATERIALS AND METHODS

Plasmid construction

p2luc plasmids

Experimental and in-frame control plasmids were cloned for each HTLV-1 *pro-pol* frameshift site (WT, ΔPK1, ΔPK2, ΔPK3, PKM1,

PKM2, and PKM3). The WT frameshift site included nucleotides 2239–2331 from the NCBI Reference Sequence NC_001436.1. The PKM1-3 frameshift sites covered the same genomic range as the WT frameshift site, whereas the ΔPK1-3 frameshift sites preserved the WT sequence but covered a smaller genomic range. These deletion constructs included nucleotides 2239–2316 (ΔPK1), nucleotides 2239–2313 (ΔPK2), and nucleotides 2239–2251 and 2327–2331 (ΔPK3). In each p2luc plasmid (Grentzmann et al. 1998), the *Renilla* (*rluc*) and firefly (*fluc*) luciferase genes were separated by DNA coding for a frameshift site (Supplemental Table S1). In the experimental plasmids, *fluc* was positioned downstream from a functional frameshift site and was in the −1 reading frame relative to *rluc*. This positioning is analogous to the positioning of the *pro* and *pol* genes in the HTLV-1 genome. In the in-frame control plasmids, the frameshift site slippery sequence was modified (TTTAAAC was changed to CTTCAAC) to create a nonfunctional frameshift site and an additional nucleotide was added downstream to place *fluc* in-frame with *rluc* (Supplemental Table S1). A p2luc out-of-frame background control plasmid was also cloned. This included the nonfunctional frameshift site from the ΔPK3 in-frame control and kept *fluc* in the −1 reading frame relative to *rluc*. In all plasmids, a T7 RNA polymerase promoter was included upstream of the *rluc* gene and a PmlI restriction site was included downstream from the *fluc* gene. Cloning was accomplished for the WT, ΔPK1, and PKM1-3 plasmids as previously described (Mouzakis et al. 2013). Briefly, complementary synthetic oligonucleotides (Integrated DNA Technologies Inc.) with BamHI and SacI compatible ends (Supplemental Table S1) were annealed, phosphorylated, and ligated into a p2luc vector. This vector had been linearized with the BamHI and SacI restriction enzymes (NEB) and dephosphorylated with recombinant Shrimp Alkaline Phosphatase (NEB). The ΔPK2, ΔPK3, and out-of-frame background control plasmids were prepared by restriction digestion of gBlock Gene Fragments (Supplemental Table S1, Integrated DNA Technologies Inc.) with BamHI and SacI, purified using a QIAquick PCR cleanup kit (Qiagen), and ligated into the p2luc vector using a 1:1.5 molar ratio of vector to insert. Recombinant plasmids were transformed into 5-alpha Competent *E. coli* (NEB), purified (Qiagen), and their sequences were verified with Sanger Sequencing (University of Wisconsin-Madison Biotechnology Center and Laragen). Plasmid DNA was linearized using the PmlI restriction enzyme (NEB) and purified using QIAquick PCR cleanup kits (Qiagen).

puc19 plasmids

Two puc19 plasmids, WT and PKM3, were cloned. To prepare the DNA inserts for cloning, gBlock Gene Fragments (Supplemental Table S1, Integrated DNA Technologies Inc.) were digested with the BamHI and EcoRI restriction enzymes (NEB). The digested DNA was purified using either a phenol/chloroform extraction and ethanol precipitation or using a QIAquick PCR cleanup kit (Qiagen). These samples were ligated into a similarly digested and dephosphorylated puc19 vector using a 1:1.5 molar ratio of vector to insert. The recombinant DNA plasmids were transformed into 5-alpha Competent *E. coli* (NEB) and purified (Qiagen). The sequences of all plasmids were verified with Sanger Sequencing (University of Wisconsin-Madison Biotechnology Center).

PCR preparation of transcription templates

Double-stranded DNA (dsDNA) templates for the WT and PKM3 transcriptions were prepared by PCR amplification of a puc19 plasmid using two oligonucleotide primers (Supplemental Table S2). Following PCR amplification, the puc19 plasmids were digested with DpnI (NEB). dsDNA templates for the ΔPK1, PKM1, and PKM2 transcriptions were prepared by PCR assembly using four oligonucleotide primers (Supplemental Table S2), as previously described (Cordero et al. 2014a; Tian et al. 2014). From the 5′ to 3′ direction, all dsDNA templates included the T7 RNA polymerase promoter (TTCTAATACGACTCACTATA), a 5-nt buffering region, a 5′ referencing hairpin, the HTLV-1 sequence of interest, a 3′ referencing hairpin, a 3-nt buffering region, and a 21-nt primer binding site (AAAAGAAACAACAACAACAAC). The WT HTLV-1 sequence included nucleotides 2235–2331 from the NC_001436.1 reference sequence. The PKM1-3 sequences covered the same genomic range, whereas the ΔPK1 sequence only included nucleotides 2235–2316.

RNA synthesis and purification

Milligram quantities of RNA for the chemical modification experiments were in vitro transcribed using the dsDNA templates prepared by PCR (Supplemental Table S3). The 2.5 mL transcription reactions contained: dsDNA templates (0.04–0.06 μg/μL), NTPs (10 mM), transcription buffer (30 mM Tris pH 8.0, 10 mM DTT, 2 mM spermidine trihydrochloride, 0.01% Triton X-100), $MgCl_2$ (40 mM), RNasin Plus RNase A inhibitor (0.016 U/μL), His$_6$-tagged T7 RNA Polymerase (1×), and diethylpyrocarbonate (DEPC)-treated (0.002%) milli-Q water. Transcription reactions were incubated overnight at 37°C. Inorganic pyrophosphate was removed by centrifugation (10 min, 3560 rpm, 4°C). The RNA was ethanol precipitated and pelleted (30 min, 3560 rpm, 4°C). After air-drying the RNA pellet for 15 min, the RNA was resuspended in 5 mL of loading dye and purified by 8% denaturing PAGE. The RNA was identified by ultraviolet absorbance, excised from the gel, and recovered by diffusion into ~100 mL of DEPC-treated (0.002%) milli-Q water over the course of ~16 h at 4°C. The RNA was separated from the gel using a 0.22 μm Millipore Stericup filter and concentrated using Amicon Ultra-15 (30k MWCO) centrifugal filters. Once the RNA reached a volume of ~300 μL, an additional 45 mL of DEPC-treated milli-Q water was centrifuged through the filter unit. RNA integrity was checked with 8 or 10% denaturing PAGE.

Microgram quantities of RNA for the frameshift assay were in vitro transcribed and purified with minor differences from a previously described protocol (Mouzakis et al. 2013). The 200 μL transcription reactions included: NTPs (11.25 mM), linearized p2luc plasmid DNA (0.05–0.10 μg/μL), transcription buffer, $MgCl_2$ (38 mM), His$_6$-tagged T7 RNA Polymerase (10×), RNasin Plus RNase A inhibitor (1 U/μL), and DEPC-treated (0.002%) milli-Q water. The reactions were incubated at 37°C for 90 min. Inorganic pyrophosphate was removed by centrifugation (7.5 min, 13,000 rpm, 4°C). The template DNA was then digested with RQ1 RNase-free DNase (Promega) during a 30 min 37°C incubation. Once again, the inorganic pyrophosphate was removed by centrifugation. EDTA (8 mM final) was added to the sample to prevent degradation. The nucleic acids were phenol/chloroform extracted

and then purified with size-exclusion chromatography, which utilized a 10-mL Bio-Scale Mini Bio-Gel P-6 Desalting Cartridge (Bio-Rad). The full-length RNA fractions were combined and refolded. Refolding involved an incubation in boiling water for 5 min followed by rapid cooling on ice for 30 min. Finally, each RNA was concentrated to ~1 μg/μL using Amicon Ultra 4-mL 10 kDa MWCO centrifugal filters. RNA integrity and folding were assessed using 1.0% agarose gel electrophoresis.

In vitro frameshift assay

In vitro frameshift assays were completed as previously described (Mouzakis et al. 2013). Minor differences in the setup of the translation reactions are described here. Each 87.5 μL reaction included RNA (0.1 μg/μL), nuclease-treated Rabbit Reticulocyte Lysate (70% of the total volume, Promega), amino acids (0.02 mM, Promega), RNasin Plus RNase A inhibitor (0.8 U/μL, Promega), and DEPC-treated (0.002%) milli-Q water. Each reaction was incubated at 37°C for 90 min and quenched with the addition of 0.5 μL of 0.156 M EDTA (0.9 mM final). Aliquots of 12.5 μL were transferred to six nonadjacent wells in a 96-Well Half-Area Polystyrene plate (Corning). These six wells represent six technical replicates associated with each biological replicate. Luminescence was measured using a dual-luciferase reporter assay (Promega) and a GloMax Navigator dual-injector Microplate Luminometer (Promega), as previously described (Marcheschi et al. 2007; Mouzakis et al. 2013). Following the addition of 25 μL of the Luciferase Assay Reagent II (Promega) and a 2 sec delay, firefly luminescence was measured for 10 sec. *Renilla* luminescence was measured after an equivalent addition of the Stop & Glo (Promega) reagent and time delay.

Frameshift efficiencies were calculated as follows. For each biological replicate for a given RNA reporter, the average ratio of firefly/*Renilla* luminescence was calculated using the firefly/*Renilla* luminescence ratios from four to six technical replicates. Data sets were analyzed using a previously published bicistronic reporter assay analysis pipeline (Supplemental Table S5; Jacobs and Dinman 2004). A freely available online tutorial (https://docs.wixstatic.com/ugd/afc9ec_a0be38bda194426ab10887e390deb720.pdf) was followed. This pipeline permitted the identification of outliers, verification that the data was normally distributed, calculation of the number of replicates required for sound statistical analysis, and propagation of error in frameshift efficiency calculations. Biological replicates were repeated until the number of replicates required for statistical analysis was met or exceeded. The minimum corrected number of replicates (Kupper and Hafner 1989) was calculated with 95% confidence in the sample mean and limited the acceptable error to 10%. For each frameshift site with experimental and in-frame control reporter RNAs (WT, ΔPK1-3, PKM1-3, and HIV-1) the frameshift efficiency was calculated by dividing the firefly/*Renilla* luminescence ratio of the experimental RNA by its corresponding in-frame control (Supplemental Equation S1). In contrast, the out-of-frame background control frameshift efficiency was not normalized to an in-frame control. Frameshift efficiency variance, standard deviation, and standard error (Supplemental Equation S2) were calculated as previously outlined (Jacobs and Dinman 2004). Unpaired two-sample *t*-tests at a 95% confidence level were used to determine if the frameshift efficiencies measured were statistically different from the WT frameshift efficiency. Here, the *P*-value reflects the probability of an incorrect conclusion. A data summary for each RNA reporter is shown in Supplemental Table S5.

RNA refolding

Folding conditions were optimized to ensure that the RNA was monomeric prior to chemical probing. Initial conditions tested were derived from a previously published protocol (Cordero et al. 2014b). In this protocol, RNA (0.08 μM) was mixed with refolding buffer (50 mM Na-HEPES, pH 8.0, 10 mM MgCl$_2$) and incubated at room temperature for 20 min. Starting with these parameters, the impact of changes in MgCl$_2$ concentration (0, 1, 5, 10, 13.3 mM) and incubation temperature (4°C, room temperature, 37°C) were examined. For RNAs prone to multimerization in the presence of MgCl$_2$ (ΔPK1, PKM1, and PKM2), the addition of a denaturation and rapid cooling sequence prior to refolding was key. For the ΔPK1, PKM1, and PKM2 RNAs, denaturation and cooling parameters were varied until monomeric conditions were empirically determined. Variables included: denaturation time (2 or 5 min at 90°C), cooling time (2, 5, or 10 min on ice), RNA concentration (0.08, 0.10, 0.11, 1.2, or 0.16 μM), and the inclusion of buffer (no buffer or 50 mM Na-HEPES pH 8.0 with MgCl$_2$ [0, 5, or 10 mM]). Given the dilute concentrations of the RNA after refolding (0.08 μM), RNAs were spin-concentrated using Amicon-Ultra (0.5 mL, 10 kDa MWCO) Centrifugal Filters to ~2.4 μM prior to their analysis with 6% nondenaturing PAGE. Dimer formation was not observed in the WT, PKM1, and PKM3 samples. Extensive refolding optimization was necessary for the ΔPK1 and PKM2 RNAs. While monomeric refolding was possible in a magnesium-free buffer (50 mM Na-HEPES, pH 8.0), each RNA formed extensive intramolecular interactions when relatively high concentrations of MgCl$_2$ (5 and 10 mM) were included, as assessed by nondenaturing PAGE following spin concentration. To limit these intramolecular interactions, the ΔPK1 and PKM2 RNAs were refolded in a lower concentration of MgCl$_2$ (1 mM). A mixture of monomer and dimer were observed under these conditions (Supplemental Fig. S7).

Each RNA was refolded under optimized conditions as follows. The WT and PKM3 RNAs (0.08 μM) were incubated in folding buffer (10 mM MgCl$_2$, 50 mM Na-HEPES, pH 8.0) for 20 min at 37°C. The PKM1 RNA (0.16 μM) was incubated for 5 min at 90°C, cooled on ice for 5 min, diluted with an equal volume of 2X folding buffer (20 mM MgCl$_2$, 100 mM Na-HEPES, pH 8.0), and incubated for 20 min at 37°C. The ΔPK1 and PKM2 RNAs (0.08 μM, 600 μL) were incubated for 5 min at 90°C in a MgCl$_2$-free buffer (50 mM Na-HEPES, pH 8.0) and then cooled on ice for 10 min. Magnesium chloride (0.6 μL, 1 M MgCl$_2$) was added and the samples (0.08 μM RNA, 50 mM Na-HEPES, pH 8.0, 1 mM MgCl$_2$) were incubated for 20 min at room temperature.

Chemical modification

Chemical modification experiments were carried out in 96-well format as previously described (Cordero et al. 2014a; Tian et al. 2014; Hartwick et al. 2018). The refolded RNAs (1.2 pmol in 15 μL) were modified by adding 5 μL of freshly prepared modification reagent (12 mg/mL N-methylisatoic anhydride [NMIA] or 1% dimethyl

sulfate [DMS]). The modification reagents were prepared by mixing: 12 mg of NMIA with 1.0 mL of anhydrous dimethyl sulfoxide (DMSO) and 10 µL of dimethyl sulfate with 90 µL of 100% ethanol and 900 µL of filter-sterilized (0.22 µm) milli-Q water. For all RNAs probed, anhydrous DMSO and 9% ethanol (or 10% for WT) were used as no modification background controls for the NMIA and DMS experiments, respectively. Modification reactions were incubated at room temperature for 15 (DMS) or 30 min (NMIA).

The 20 µL reactions were quenched with 9.75 µL of a bead-quench reagent and incubated at room temperature for 7 min. The bead-quench reagent included: 1.5 µL of clean oligo-dT magnetic beads (Ambion), 0.25 µL of 0.25 µM primer (6-FAM-A$_{20}$-GTTGTTGTTGTTGTTTCTTTT), 3 µL of 5M NaCl, and 5 µL of quench reagent (0.5 M Na-MES, pH 6.0 for NMIA; 2-mercaptoethanol for DMS). Following quenching and primer binding, the 96-well plate was placed on a 96-post magnetic-stand (VP Scientific) and set at room temperature for 10 min. The samples were washed twice with 100 µL of 70% ethanol, air-dried for 10 min, and resuspended in 2.5 µL of DEPC-treated (0.002%) milli-Q water. After removing the 96-well plate from the magnetic stand, each sample received 2.5 µL of Superscript III Reverse Transcription (RT) mix. The RT mix included 1 µL of 5X First Strand buffer (Fisher Scientific), 0.25 µL of 0.1 M DTT, 0.4 µL of 10 mM dNTPs, 0.75 µL of DEPC-treated (0.002%) milli-Q water, and 0.1 µL of Superscript III Reverse Transcriptase (200 U/µL, Fisher Scientific). The sample plate was incubated for 30 min at 42°C. RNAs were hydrolyzed with the addition of 5 µL of 0.4 M NaOH and a 3 min incubation at 90°C. The sample plate was cooled on ice for 3 min and the reactions were neutralized with 5 µL of an acid-quench mixture (1.4 M NaCl, 0.6 M HCl, and 1.3 M NaOAc).

FAM-labeled complementary DNA (cDNA) was purified and separated using the 96-post magnetic stand. After leaving the sample plate on the magnetic stand for 7 min, the samples were washed twice with 100 µL of 70% ethanol and air-dried for 10 min. The cDNA was eluted in 11 µL of a ROX-formamide mixture, which was prepared by mixing 1.375 µL of ROX 350 ladder (Applied Biosystems) with 600 µL of HiDi-formamide (Applied Biosystems). After a 15 min elution, the cDNA samples were transferred to a capillary electrophoresis (CE) optical plate and analyzed with an ABI 3500 Genetic Analyzer Capillary Electrophoresis instrument.

Chemical modification experiments with NMIA and DMS were repeated a minimum of three times. After their initial analysis with CE, all replicates were subject to a 10-fold dilution (or a 20-fold dilution for the PKM3 NMIA samples) in the ROX-formamide mixture and analyzed again. Sequencing ladders for each RNA were prepared by reverse transcription in the presence of dideoxynucleoside triphosphates (ddNTPs), as previously described (Cordero et al. 2014a). These reactions were analogous to the reverse transcription reactions for the chemical modification experiments, with the exception of the RT mix composition, which included equimolar amounts of 1 mM dNTPs and ddNTPs.

Data processing, structural modeling, and visualization

The HiTRACE software package and a MATLAB toolbox (freely available at https://ribokit.github.io/HiTRACE/) were used to quantify and normalize nucleotide reactivities, as previously described (Kladwang et al. 2011a; Yoon et al. 2011; Kim et al. 2013; Kladwang et al. 2014; Tian et al. 2014; Lee et al. 2015; Hartwick et al. 2018). Briefly, the CE traces were aligned. Their sequence assignments were completed manually and verified using the sequencing ladders. The CE traces were corrected for signal attenuation, background subtracted, and then normalized using the reactivities of the 5′ and 3′ reference hairpin loop nucleotides (GAGUA). The nucleotide reactivities were quantified with a Gaussian fit and averaged. Error estimates were calculated using HiTRACE and points with more scatter than expected were deemed to be outliers. Outliers were identified only in the WT DMS data set.

Data-driven secondary structures were obtained using the Shapeknots program of the RNAstructure package (version 5.6) (Reuter and Mathews 2010). The full-length RNA sequences (Supplemental Table S4) were used in combination with the normalized and averaged NMIA nucleotide reactivity data to generate all secondary structures. Bootstrapping methods were used to extrapolate the dominant secondary structure and to calculate the structure's helix-wide confidence values (Kladwang et al. 2011b; Tian et al. 2014). The latter describe the likelihood of individual helix formation (Kladwang et al. 2011b). NMIA nucleotide reactivities were superimposed onto the secondary structures and visualized with the VARNA program (Darty et al. 2009). The normalized and averaged DMS nucleotide reactivities were used to validate these NMIA-derived secondary structures. Specifically, the A and C DMS data were overlaid onto the NMIA-derived secondary structure and visualized with VARNA. Prior to their use in VARNA (Darty et al. 2009), all outlier nucleotide reactivities were set to zero. Varna displayed the nucleotide reactivities with a colored circle as follows: Nonreactive nucleotides (those with reactivities ≤ 10% of the normalized maximum) were white, moderately reactive nucleotides (those with reactivities within 10%–80% of the normalized maximum) were displayed as a gradient, and highly reactive nucleotides (those with reactivities ≥ 80% of the normalized maximum) were dark red (NMIA) or dark blue (DMS). The authors note that 2 bp predicted in the loops of ΔPK1 and PKM1 between nucleotides A2268–U2276 and C2269–G2275 were not included in the reported figures (Fig. 5A,B) because it is generally accepted that it takes a minimum of 3 bp to form a stable A-form helix.

Sequence conservation and covariation analysis

The sequence conservation and base-pair covariation of the HTLV-1 *pro-pol* frameshift site was assessed using the HTLV-1 sequences available (as of January 13, 2019) in the NCBI database. Specifically, NCBI's Standard Nucleotide Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/Blast.cgi) was used to search for more dissimilar sequences to the 2239-2324 query subrange of the NC_001436.1 sequence. All bovine, simian, and other HTLV subtypes (2, 3, and 4) sequence results were excluded from this search set. The maximum number of target sequences was set to 1000. Three hundred and fifty-two unique HTLV-1 sequences were identified. These were aligned in Ugene (Okonechnikov et al. 2012) allowing for mismatches. The aligned file was exported as a .ugenedb file and manually converted to a Stockholm file. Base-pair covariation was assessed

with R-Scape (Rivas et al. 2017) and sequence conservation with UGene (Okonechnikov et al. 2012).

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

REFERENCES

Atkins JF, Loughran G, Bhatt PR, Firth AE, Baranov PV. 2016. Ribosomal frameshifting and transcriptional slippage: from genetic steganography and cryptography to adventitious use. *Nucleic Acids Res* **44:** 7007–7078. doi:10.1093/nar/gkw530

Bangham CRM. 2018. Human T cell leukemia virus type 1: persistence and pathogenesis. *Annu Rev Immunol* **36:** 43–71. doi:10.1146/annurev-immunol-042617-053222

Bangham CRM, Araujo A, Yamano Y, Taylor GP. 2015. HTLV-1-associated myelopathy/tropical spastic paraparesis. *Nat Rev Dis Primers* **1:** 15012. doi:10.1038/nrdp.2015.12

Biswas P, Jiang X, Pacchia AL, Dougherty JP, Peltz SW. 2004. The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* **78:** 2082–2087. doi:10.1128/JVI.78.4.2082-2087.2004

Brierley I. 1995. Ribosomal frameshifting viral RNAs. *J Gen Virol* **76:** 1885–1892. doi:10.1099/0022-1317-76-8-1885

Brierley I, Dos Ramos FJ. 2006. Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res* **119:** 29–42. doi:10.1016/j.virusres.2005.10.008

Brierley I, Digard P, Inglis SC. 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. *Cell* **57:** 537–547. doi:10.1016/0092-8674(89)90124-4

Brierley I, Jenner AJ, Inglis SC. 1992. Mutational analysis of the "slippery-sequence" component of a coronavirus ribosomal frameshifting signal. *J Mol Biol* **227:** 463–479. doi:10.1016/0022-2836(92)90901-U

Brierley I, Gilbert RJ, Pennell S. 2008. RNA pseudoknots and the regulation of protein synthesis. *Biochem Soc Trans* **36:** 684–689. doi:10.1042/BST0360684

Brierley I, Gilbert RJC, Pennell S. 2010. Pseudoknot-dependent programmed −1 ribosomal frameshifting: structures, mechanisms and models. In *Recoding: expansion of decoding rules enriches gene expression* (ed. Atkins JF, Gesteland RF), Vol. 24, pp. 149–174. Springer, New York.

Butcher SE, Pyle AM. 2011. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc Chem Res* **44:** 1302–1311. doi:10.1021/ar200098t

Caliskan N, Katunin VI, Belardinelli R, Peske F, Rodnina MV. 2014. Programmed −1 frameshifting by kinetic partitioning during impeded translocation. *Cell* **157:** 1619–1631. doi:10.1016/j.cell.2014.04.041

Caliskan N, Peske F, Rodnina MV. 2015. Changed in translation: mRNA recoding by −1 programmed ribosomal frameshifting. *Trends Biochem Sci* **40:** 265–274. doi:10.1016/j.tibs.2015.03.006

Carpentier A, Barez PY, Hamaidia M, Gazon H, de Brogniez A, Perike S, Gillet N, Willems L. 2015. Modes of human T cell leukemia virus type 1 transmission, replication and persistence. *Viruses* **7:** 3603–3624. doi:10.3390/v7072793

Chamorro M, Parkin N, Varmus HE. 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. *Proc Natl Acad Sci* **89:** 713–717. doi:10.1073/pnas.89.2.713

Chapman EG, Moon SL, Wilusz J, Kieft JS. 2014. RNA structures that resist degradation by Xrn1 produce a pathogenic Dengue virus RNA. *Elife* **3:** e01892. doi:10.7554/eLife.01892

Charbonneau J, Gendron K, Ferbeyre G, Brakier-Gingras L. 2012. The 5′ UTR of HIV-1 full-length mRNA and the Tat viral protein modulate the programmed −1 ribosomal frameshift that generates HIV-1 enzymes. *RNA* **18:** 519–529. doi:10.1261/rna.030346.111

Chen G, Chang KY, Chou MY, Bustamante C, Tinoco I Jr. 2009. Triplex structures in an RNA pseudoknot enhance mechanical stability and increase efficiency of −1 ribosomal frameshifting. *Proc Natl Acad Sci* **106:** 12706–12711. doi:10.1073/pnas.0905046106

Chen YT, Chang KC, Hu HT, Chen YL, Lin YH, Hsu CF, Chang CF, Chang KY, Wen JD. 2017. Coordination among tertiary base pairs results in an efficient frameshift-stimulating RNA pseudoknot. *Nucleic Acids Res* **45:** 6011–6022. doi:10.1093/nar/gkx134

Cordero P, Kladwang W, VanLang CC, Das R. 2014a. The mutate-and-map protocol for inferring base pairs in structured RNA. In *RNA folding: methods and protocols* (ed. Waldisch C), Vol. 1086, pp. 53–77. Humana Press, NJ.

Cordero P, Kladwang W, VanLang CC, Das R. 2014b. The mutate-and-map protocol for inferring base pairs in structured RNA. *Methods Mol Biol* **1086:** 53–77. doi:10.1007/978-1-62703-667-2_4

Cornish PV, Hennig M, Giedroc DP. 2005. A loop 2 cytidine-stem 1 minor groove interaction as a positive determinant for pseudoknot-stimulated −1 ribosomal frameshifting. *Proc Natl Acad Sci* **102:** 12694–12699. doi:10.1073/pnas.0506166102

Cornish PV, Giedroc DP, Hennig M. 2006. Dissecting non-canonical interactions in frameshift-stimulating mRNA pseudoknots. *J Biomol NMR* **35:** 209–223. doi:10.1007/s10858-006-9033-x

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* **25:** 1974–1975. doi:10.1093/bioinformatics/btp250

Demeshkina N, Jenner L, Westhof E, Yusupov M, Yusupova G. 2012. A new understanding of the decoding principle on the ribosome. *Nature* **484:** 256–259. doi:10.1038/nature10913

Devi G, Zhou Y, Zhong Z, Toh D-FK, Chen G. 2015. RNA triplexes: from structural principles to biological and biotech applications. *Wiley Interdiscip Rev RNA* **6:** 111–128. doi:10.1002/wrna.1261

Dinman JD. 2012a. Control of gene expression by translational recoding. *Adv Protein Chem Struct Biol* **86:** 129–149. doi:10.1016/B978-0-12-386497-0.00004-9

Dinman JD. 2012b. Mechanisms and implications of programmed translational frameshifting. *Wiley Interdiscip Rev RNA* **3:** 661–673. doi:10.1002/wrna.1126

Ehresmann C, Baudin F, Mougel M, Romby P, Ebel JP, Ehresmann B. 1987. Probing the structure of RNAs in solution. *Nucleic Acids Res* **15:** 9109–9128. doi:10.1093/nar/15.22.9109

Fan N, Gavalchin J, Paul B, Wells KH, Lane MJ, Poiesz BJ. 1992. Infection of peripheral blood mononuclear cells and cell lines by cell-free human T-cell lymphoma/leukemia virus type I. *J Clin Microbiol* **30:** 905–910. doi:10.1128/JCM.30.4.905-910.1992

Farabaugh PJ. 1996. Programmed translational frameshifting. *Microbiol Rev* **60:** 103–134. doi:10.1128/MMBR.60.1.103-134.1996

Futsch N, Mahieux R, Dutartre H. 2017. HTLV-1, the other pathogenic yet neglected human retrovirus: from transmission to therapeutic treatment. *Viruses* **10:** 1–25. doi:10.3390/v10010001

Garcia-Miranda P, Becker JT, Benner BE, Blume A, Sherer NM, Butcher SE. 2016. Stability of HIV frameshift site RNA correlates with frameshift efficiency and decreased virus infectivity. *J Virol* **90:** 6906–6917. doi:10.1128/JVI.00149-16

Gendron K, Charbonneau J, Dulude D, Heveker N, Ferbeyre G, Brakier-Gingras L. 2008. The presence of the TAR RNA structure alters the programmed −1 ribosomal frameshift efficiency of the human immunodeficiency virus type 1 (HIV-1) by modifying the rate of translation initiation. *Nucleic Acids Res* **36:** 30–40. doi:10.1093/nar/gkm906

Gessain A, Cassar O. 2012. Epidemiological aspects and world distribution of HTLV-1 infection. *Front Microbiol* **3:** 1–23. doi:10.3389/fmicb.2012.00388

Gessain A, Barin F, Vernant JC, Gout O, Maurs L, Calender A, de Thé G. 1985. Antibodies to human T-lymphotropic virus type-I in patients with tropical spastic paraparesis. *Lancet* **2:** 407–410. doi:10.1016/S0140-6736(85)92734-5

Gessain A, Gallo RC, Franchini G. 1992. Low degree of human T-cell leukemia/lymphoma virus type I genetic drift in vivo as a means of monitoring viral transmission and movement of ancient human populations. *J Virol* **66:** 2288–2295. doi:10.1128/JVI.66.4.2288-2295.1992

Gesteland RF, Atkins JF. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65:** 741–768. doi:10.1146/annurev.bi.65.070196.003521

Giedroc DP, Cornish PV. 2009. Frameshifting RNA pseudoknots: structure and mechanism. *Virus Res* **139:** 193–208. doi:10.1016/j.virusres.2008.06.008

Giedroc DP, Theimer CA, Nixon PL. 2000. Structure, stability and function of RNA pseudoknots involved in stimulating ribosomal frameshifting. *J Mol Biol* **298:** 167–185. doi:10.1006/jmbi.2000.3668

Grentzmann G, Ingram JA, Kelly PJ, Gesteland RF, Atkins JF. 1998. A dual-luciferase reporter system for studying recoding signals. *RNA* **4:** 479–486. doi:10.1017/S1355838298971576

Gupta A, Bansal M. 2014. Local structural and environmental factors define the efficiency of an RNA pseudoknot involved in programmed ribosomal frameshift process. *J Phys Chem B* **118:** 11905–11920. doi:10.1021/jp507154u

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

Halma MTJ, Ritchie DB, Cappellano TR, Neupane K, Woodside MT. 2019. Complex dynamics under tension in a high-efficiency frameshift stimulatory structure. *Proc Natl Acad Sci* **116:** 19500–19505. doi:10.1073/pnas.1905258116

Hartwick EW, Costantino DA, MacFadden A, Nix JC, Tian S, Das R, Kieft JS. 2018. Ribosome-induced RNA conformational changes in a viral 3′-UTR sense and regulate translation levels. *Nat Commun* **9:** 5074. doi:10.1038/s41467-018-07542-x

Honigman A, Falk H, Mador N, Rosental T, Panet A. 1995. Translation efficiency of the human T-cell leukemia virus (HTLV-2) gag gene modulates the frequency of ribosomal frameshifting. *Virology* **208:** 312–318. doi:10.1006/viro.1995.1154

Huang X, Cheng Q, Du Z. 2013. A genome-wide analysis of RNA pseudoknots that stimulate efficient −1 ribosomal frameshifting or readthrough in animal viruses. *BioMed Res Int* **2013:** 1–15. doi:10.1155/2013/984028

Hung M, Patel P, Davis S, Green SR. 1998. Importance of ribosomal frameshifting for human immunodeficiency virus type 1 particle assembly and replication. *J Virol* **72:** 4819–4824. doi:10.1128/JVI.72.6.4819-4824.1998

Jacobs JL, Dinman JD. 2004. Systematic analysis of bicistronic reporter assay data. *Nucleic Acids Res* **32:** e160. doi:10.1093/nar/gnh157

Jones KS, Petrow-Sadowski C, Huang YK, Bertolette DC, Ruscetti FW. 2008. Cell-free HTLV-1 infects dendritic cells leading to transmission and transformation of CD4$^+$ T cells. *Nat Med* **14:** 429–436. doi:10.1038/nm1745

Karacostas V, Wolffe EJ, Nagashima K, Gonda MA, Moss B. 1993. Overexpression of the HIV-1 gag-pol polyprotein results in intracellular activation of HIV-1 protease and inhibition of assembly and budding of virus-like particles. *Virology* **193:** 661–671. doi:10.1006/viro.1993.1174

Kim HK, Tinoco I Jr. 2017. EF-G catalyzed translocation dynamics in the presence of ribosomal frameshifting stimulatory signals. *Nucleic Acids Res* **45:** 2865–2874. doi:10.1093/nar/gkw1020

Kim YG, Maas S, Rich A. 2001. Comparative mutational analysis of cis-acting RNA signals for translational frameshifting in HIV-1 and HTLV-2. *Nucleic Acids Res* **29:** 1125–1131. doi:10.1093/nar/29.5.1125

Kim H, Cordero P, Das R, Yoon S. 2013. HiTRACE-Web: an online tool for robust analysis of high-throughput capillary electrophoresis. *Nucleic Acids Res* **41:** W492–W498. doi:10.1093/nar/gkt501

Kladwang W, Cordero P, Das R. 2011a. A mutate-and-map strategy accurately infers the base pairs of a 35-nucleotide model RNA. *RNA* **17:** 522–534. doi:10.1261/rna.2516311

Kladwang W, VanLang CC, Cordero P, Das R. 2011b. A two-dimensional mutate-and-map strategy for non-coding RNA structure. *Nat Chem* **3:** 954–962. doi:10.1038/nchem.1176

Kladwang W, Mann TH, Becka A, Tian S, Kim H, Yoon S, Das R. 2014. Standardization of RNA chemical mapping experiments. *Biochemistry* **53:** 3063–3065. doi:10.1021/bi5003426

Kontos H, Napthine S, Brierley I. 2001. Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* **21:** 8657–8670. doi:10.1128/MCB.21.24.8657-8670.2001

Korniy N, Samatova E, Anokhina MM, Peske F, Rodnina MV. 2019. Mechanisms and biomedical implications of −1 programmed

ribosome frameshifting on viral and bacterial mRNAs. *FEBS Lett* **593:** 1468–1482. doi:10.1002/1873-3468.13478

Krump NA, You J. 2018. Molecular mechanisms of viral oncogenesis in humans. *Nat Rev Microbiol* **16:** 684–698. doi:10.1038/s41579-018-0064-6

Kupper LL, Hafner KB. 1989. How appropriate are popular sample size formulas? *Am Stat* **43:** 101–105. doi:10.1080/00031305.1989.10475628

Kurland CG. 1992. Translational accuracy and the fitness of bacteria. *Annu Rev Genet* **26:** 29–50. doi:10.1146/annurev.ge.26.120192.000333

Le SY, Shapiro BA, Chen JH, Nussinov R, Maizel JV. 1991. RNA pseudoknots downstream of the frameshift sites of retroviruses. *Genet Anal Tech Appl* **8:** 191–205. doi:10.1016/1050-3862(91)90013-H

Lee S, Kim H, Tian S, Lee T, Yoon S, Das R. 2015. Automated band annotation for RNA structure probing experiments with numerous capillary electrophoresis profiles. *Bioinformatics* **31:** 2808–2815. doi:10.1093/bioinformatics/btv282

Leonard CW, Hajdin CE, Karabiber F, Mathews DH, Favorov OV, Dokholyan NV, Weeks KM. 2013. Principles for understanding the accuracy of SHAPE-directed RNA structure modeling. *Biochemistry* **52:** 588–595. doi:10.1021/bi300755u

Liao P-Y, Choi YS, Dinman JD, Lee KH. 2011. The many paths to frameshifting: kinetic modelling and analysis of the effects of different elongation steps on programmed –1 ribosomal frameshifting. *Nucleic Acids Res* **39:** 300–312. doi:10.1093/nar/gkq761

Lopinski JD, Dinman JD, Bruenn JA. 2000. Kinetics of ribosomal pausing during programmed –1 translational frameshifting. *Mol Cell Biol* **20:** 1095–1103. doi:10.1128/MCB.20.4.1095-1103.2000

Low JT, Garcia-Miranda P, Mouzakis KD, Gorelick RJ, Butcher SE, Weeks KM. 2014. Structure and dynamics of the HIV-1 frameshift element RNA. *Biochemistry* **53:** 4282–4291. doi:10.1021/bi5004926

Mador N, Panet A, Honigman A. 1989. Translation of *gag, pro*, and *pol* gene products of human T-cell leukemia virus type 2. *J Virol* **63:** 2400–2404. doi:10.1128/JVI.63.5.2400-2404.1989

Marcheschi RJ, Staple DW, Butcher SE. 2007. Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem–loop. *J Mol Biol* **373:** 652–663. doi:10.1016/j.jmb.2007.08.033

Marcheschi RJ, Mouzakis KD, Butcher SE. 2009. Selection and characterization of small molecules that bind the HIV-1 frameshift site RNA. *ACS Chem Biol* **4:** 844–854. doi:10.1021/cb900167m

Martin JL, Maldonado JO, Mueller JD, Zhang W, Mansky LM. 2016. Molecular studies of HTLV-1 replication: an update. *Viruses* **8:** 31. doi:10.3390/v8020031

Martin F, Tagaya Y, Gallo R. 2018. Time to eradicate HTLV-1: an open letter to WHO. *Lancet* **391:** 1893–1894. doi:10.1016/S0140-6736(18)30974-7

Matsuoka M, Jeang KT. 2007. Human T-cell leukaemia virus type 1 (HTLV-1) infectivity and cellular transformation. *Nat Rev Cancer* **7:** 270–280. doi:10.1038/nrc2111

Meissner ME, Mendonça LM, Zhang W, Mansky LM. 2017. Polymorphic nature of human T-cell leukemia virus type 1 particle cores as revealed through characterization of a chronically infected cell line. *J Virol* **91:** e00369-17. doi:10.1128/JVI.00369-17

Merino EJ, Wilkinson KA, Coughlan JL, Weeks KM. 2005. RNA structure analysis at single nucleotide resolution by selective 2′-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127:** 4223–4231. doi:10.1021/ja043822v

Meydan S, Klepacki D, Karthikeyan S, Margus T, Thomas P, Jones JE, Khan Y, Briggs J, Dinman JD, Vázquez-Laslop N, et al. 2017. Programmed ribosomal frameshifting generates a copper transporter and a copper chaperone from the same gene. *Mol Cell* **65:** 207–219. doi:10.1016/j.molcel.2016.12.008

Michiels PJ, Versleijen AA, Verlaan PW, Pleij CW, Hilbers CW, Heus HA. 2001. Solution structure of the pseudoknot of SRV-1 RNA, involved in ribosomal frameshifting. *J Mol Biol* **310:** 1109–1123. doi:10.1006/jmbi.2001.4823

Mouzakis KD, Lang AL, Vander Meulen KA, Easterday PD, Butcher SE. 2013. HIV-1 frameshift efficiency is primarily determined by the stability of base pairs positioned at the mRNA entrance channel of the ribosome. *Nucleic Acids Res* **41:** 1901–1913. doi:10.1093/nar/gks1254

Nam SH, Copeland TD, Hatanaka M, Oroszlan S. 1993. Characterization of ribosomal frameshifting for expression of pol gene products of human T-cell leukemia virus type I. *J Virol* **67:** 196–203. doi:10.1128/JVI.67.1.196-203.1993

Napthine S, Liphardt J, Bloys A, Routledge S, Brierley I. 1999. The role of RNA pseudoknot stem 1 length in the promotion of efficient -1 ribosomal frameshifting. *J Mol Biol* **288:** 305–320. doi:10.1006/jmbi.1999.2688

Okonechnikov K, Golosova O, Fursov M; UGENE team. 2012. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics* **28:** 1166–1167. doi:10.1093/bioinformatics/bts091

Olsthoorn RCL, Reumerman R, Hilbers CW, Pleij CWA, Heus HA. 2010. Functional analysis of the SRV-1 RNA frameshifting pseudoknot. *Nucleic Acids Res* **38:** 7665–7672. doi:10.1093/nar/gkq629

Osame M, Usuku K, Izumo S, Ijichi N, Amitani H, Igata A, Matsumoto M, Tara M. 1986. HTLV-I associated myelopathy, a new clinical entity. *Lancet* **1:** 1031–1032. doi:10.1016/S0140-6736(86)91298-5

Park J, Morrow CD. 1991. Overexpression of the *gag-pol* precursor from human immunodeficiency virus type 1 proviral genomes results in efficient proteolytic processing in the absence of virion production. *J Virol* **65:** 5111–5117. doi:10.1128/JVI.65.9.5111-5117.1991

Peattie DA, Gilbert W. 1980. Chemical probes for higher-order structure in RNA. *Proc Natl Acad Sci* **77:** 4679–4682. doi:10.1073/pnas.77.8.4679

Pennell S, Manktelow E, Flatt A, Kelly G, Smerdon SJ, Brierley I. 2008. The stimulatory RNA of the Visna-Maedi retrovirus ribosomal frameshifting signal is an unusual pseudoknot with an interstem element. *RNA* **14:** 1366–1377. doi:10.1261/rna.1042108

Peselis A, Serganov A. 2014. Structure and function of pseudoknots involved in gene expression control. *Wiley Interdiscip Rev RNA* **5:** 803–822. doi:10.1002/wrna.1247

Plant EP. 2003. The 9-Å solution: how mRNA pseudoknots promote efficient programmed –1 ribosomal frameshifting. *RNA* **9:** 168–174. doi:10.1261/rna.2132503

Plant EP, Rakauskaite R, Taylor DR, Dinman JD. 2010. Achieving a golden mean: mechanisms by which coronaviruses ensure synthesis of the correct stoichiometric ratios of viral proteins. *J Virol* **84:** 4330–4340. doi:10.1128/JVI.02480-09

Poiesz BJ, Ruscetti FW, Gazdar AF, Bunn PA, Minna JD, Gallo RC. 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. *Proc Natl Acad Sci* **77:** 7415–7419. doi:10.1073/pnas.77.12.7415

Reil H, Kollmus H, Weidle UH, Hauser H. 1993. A heptanucleotide sequence mediates ribosomal frameshifting in mammalian cells. *J Virol* **67:** 5579–5584. doi:10.1128/JVI.67.9.5579-5584.1993

Reuter JS, Mathews DH. 2010. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11:** 129. doi:10.1186/1471-2105-11-129

Ritchie DB, Foster DA, Woodside MT. 2012. Programmed –1 frameshifting efficiency correlates with RNA pseudoknot conformational plasticity, not resistance to mechanical unfolding. *Proc Natl Acad Sci* **109:** 16167–16172. doi:10.1073/pnas.1204114109

Rivas E, Clements J, Eddy SR. 2017. A statistical test for conserved RNA structure shows lack of evidence for structure in lncRNAs. *Nat Methods* **14:** 45–48. doi:10.1038/nmeth.4066

Rodnina MV, Korniy N, Klimova M, Karki P, Peng B-Z, Senyushkina T, Belardinelli R, Maracci C, Wohlgemuth I, Samatova E, et al. 2019. Translational recoding: canonical translation mechanisms reinterpreted. *Nucleic Acids Res* doi:10.1093/nar/gkz783. doi:10.1093/nar/gkz783

Stahl G, McCarty GP, Farabaugh PJ. 2002. Ribosome structure: revisiting the connection between translational accuracy and unconventional decoding. *Trends Biochem Sci* **27:** 178–183. doi:10.1016/S0968-0004(02)02064-9

Staple DW, Butcher SE. 2005. Pseudoknots: rNA structures with diverse functions. *PLoS Biol* **3:** 0956–0959. doi:10.1371/journal.pbio.0030213

Tagaya Y, Gallo RC. 2017. The exceptional oncogenicity of HTLV-1. *Front Microbiol* **8:** 1–4. doi:10.3389/fmicb.2017.01425

ten Dam EB, Pleij CW, Bosch L. 1990. RNA pseudoknots: translational frameshifting and readthrough on viral RNAs. *Virus Genes* **4:** 121–136. doi:10.1007/BF00678404

ten Dam EB, Verlaan PW, Pleij CW. 1995. Analysis of the role of the pseudoknot component in the SRV-1 *gag-pro* ribosomal frameshift signal: loop lengths and stability of the stem regions. *RNA* **1:** 146–154.

Tholstrup J, Oddershede LB, Sørensen MA. 2012. mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res* **40:** 303–313. doi:10.1093/nar/gkr686

Tian S, Cordero P, Kladwang W, Das R. 2014. High-throughput mutate-map-rescue evaluates SHAPE-directed RNA structure and uncovers excited states. *RNA* **20:** 1815–1826. doi:10.1261/rna.044321.114

Tijerina P, Mohr S, Russell R. 2007. DMS footprinting of structured RNAs and RNA-protein complexes. *Nat Protoc* **2:** 2608–2623. doi:10.1038/nprot.2007.380

Van Dooren S, Pybus OG, Salemi M, Liu HF, Goubau P, Remondegui C, Talarmin A, Gotuzzo E, Alcantara LCJ, Galvão-Castro B, et al. 2004. The low evolutionary rate of human T-cell lymphotropic virus type-1 confirmed by analysis of vertical transmission chains. *Mol Biol Evol* **21:** 603–611. doi:10.1093/molbev/msh053

Weeks KM. 2010. Advances in RNA structure analysis by chemical probing. *Curr Opin Struct Biol* **20:** 295–304. doi:10.1016/j.sbi.2010.04.001

Weeks KM, Mauger DM. 2011. Exploring RNA structural codes with SHAPE chemistry. *Acc Chem Res* **44:** 1280–1291. doi:10.1021/ar200051h

Wilkinson KA, Merino EJ, Weeks KM. 2006. Selective 2′-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1:** 1610–1616. doi:10.1038/nprot.2006.249

Wu B, Zhang H, Sun R, Peng S, Cooperman BS, Goldman YE, Chen C. 2018. Translocation kinetics and structural dynamics of ribosomes are modulated by the conformational plasticity of downstream pseudoknots. *Nucleic Acids Res* **46:** 9736–9748. doi:10.1093/nar/gky636

Yan S, Wen JD, Bustamante C, Tinoco I Jr. 2015. Ribosome excursions during mRNA translocation mediate broad branching of frameshift pathways. *Cell* **160:** 870–881. doi:10.1016/j.cell.2015.02.003

Yang L, Zhong Z, Tong C, Jia H, Liu Y, Chen G. 2018. Single-molecule mechanical folding and unfolding of RNA hairpins: effects of single A-U to A·C pair substitutions and single proton binding and implications for mRNA structure-induced −1 ribosomal frameshifting. *J Am Chem Soc* **140:** 8172–8184. doi:10.1021/jacs.8b02970

Yoon S, Kim J, Hum J, Kim H, Park S, Kladwang W, Das R. 2011. HiTRACE: high-throughput robust analysis for capillary electrophoresis. *Bioinformatics* **27:** 1798–1805. doi:10.1093/bioinformatics/btr277

Yoshida M, Miyoshi I, Hinuma Y. 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. *Proc Natl Acad Sci* **79:** 2031–2035. doi:10.1073/pnas.79.6.2031

Yu CH, Noteborn MH, Olsthoorn RC. 2010. Stimulation of ribosomal frameshifting by antisense LNA. *Nucleic Acids Res* **38:** 8277–8283. doi:10.1093/nar/gkq650