

# Highly conserved RNA pseudoknots at the *gag-pol* junction of HIV-1 suggest a novel mechanism of $-1$ ribosomal frameshifting

XIAOLAN HUANG,<sup>1</sup> YANG YANG,<sup>2</sup> GUAN WANG,<sup>2</sup> QIANG CHENG,<sup>1</sup> and ZHIHUA DU<sup>2,3</sup>

<sup>1</sup>Department of Computer Science, <sup>2</sup>Department of Chemistry and Biochemistry, Southern Illinois University at Carbondale, Carbondale, Illinois 62901, USA

## ABSTRACT

$-1$  programmed ribosomal frameshifting (PRF) is utilized by many viruses to synthesize their enzymatic (Pol) and structural (Gag) proteins at a defined ratio. For efficient  $-1$  PRF, two *cis*-acting elements are required: a heptanucleotide frameshift site and a downstream stimulator such as a pseudoknot. We have analyzed the *gag-pol* junction sequences from 4254 HIV-1 strains. Approximately ninety-five percent of the sequences can form four pseudoknots PK1–PK4 ( $\sim 97\%$  contain PK1, PK3, and PK4), covering  $\sim 72$  nt including the frameshift site. Some pseudoknots are mutually excluded due to sequence overlap. PK1 and PK3 arrange tandemly. Their stems form a quasi-continuous helix of  $\sim 22$  bp. We propose a novel mechanism for possible roles of these pseudoknots. Multiple alternative structures may exist at the *gag-pol* junction. In most strains, the PK1–PK3 tandem pseudoknots may dominate the structurally heterogeneous pool of RNA due to their greater overall stability. The tandem pseudoknots may function as a breaking system to slow down the ribosome. The ribosome unwinds PK1 and stem 1 of PK3 before it can reach the frameshift site. Then, PK4 can form rapidly because the intact stem 2 of PK3 makes up a large part of the stem 1 of PK4. The newly formed PK4 jams the entrance of the mRNA tunnel. The process then proceeds as in a typical case of  $-1$  PRF. This mechanism incorporates several exquisite new features while still being consistent with the current paradigm of pseudoknot-dependent  $-1$  PRF.

**Keywords:** RNA pseudoknot;  $-1$  ribosomal frameshifting; HIV-1; recoding

## INTRODUCTION

$-1$  programmed ribosomal frameshifting (PRF) is a recoding mechanism in the translation process which enables protein synthesis from two overlapping reading frames (Gesteland et al. 1992; Baranov et al. 2002).  $-1$  PRF is site-specific. It occurs at a specific heptanucleotide “slippery sequence” which has a typical composition of X XXY YYZ (XXX and YYY represent a stretch of three identical nucleotides; XXY and YYZ are two codons in the 0 reading frame). In a  $-1$  PRF event, the translating ribosome shifts back 1 nt along the mRNA at the slippery sequence. The two tRNAs in the A and P sites of the ribosome originally recognize the 0 reading frame codons XXY and YYZ; after  $-1$  PRF, the tRNAs occupy the  $-1$  reading frame codons XXX and YYY. Although  $-1$  PRF occurs at the slippery sequence, this sequence alone is often not sufficient to cause efficient frameshifting. Another *cis*-acting element in the mRNA known as a “stimulator” is required in most reported cases. The most frequently occurring

frameshift stimulator is an RNA pseudoknot located several nucleotides downstream from the slippery sequence. In some cases, the frameshift stimulator can also be a conventional stem-loop structure or some other complicated RNA structures (Pleij et al. 1985; Pleij 1990; Dam et al. 1992).

$-1$  PRF is well-established in viruses. It is utilized by a large number of viruses (including retroviruses such as HIV) to express their structural and enzymatic proteins at a defined ratio from a single mRNA (Atkins et al. 1990; Gesteland et al. 1992; Gesteland and Atkins 1996; Brierley et al. 2007). This ratio is referred to as the frameshifting efficiency, which dictates the molar ratio of viral structural and enzymatic proteins. Because maintaining such a defined ratio is of vital importance to the viral life cycle, the viral  $-1$  PRF signals (especially the stimulator RNA structures) may represent valuable targets for the development of antiviral therapeutics (Biswas et al. 2004; Gareiss and Miller 2009; Marcheschi

<sup>3</sup>Corresponding author

E-mail [zdu@chem.siu.edu](mailto:zdu@chem.siu.edu)

Article published online ahead of print. Article and publication date are at <http://www.rnajournal.org/cgi/doi/10.1261/rna.042457.113>.

© 2014 Huang et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see <http://rnajournal.cshlp.org/site/misc/terms.xhtml>). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

et al. 2011). Small molecules that bind to the stimulator RNA structures and modulate the frameshifting efficiency of the retrovirus HIV-1 and the coronavirus SARS (SARS CoV) were identified (McNaughton et al. 2007; Marcheschi et al. 2009; Park et al. 2011).

In most viruses known or expected to use the  $-1$  PRF mechanism, the frameshift stimulator RNA structure is confirmed or expected to be an H (hairpin)-type pseudoknot, separated from the slippery sequence by a spacer region of several nucleotides (Brierley 1995; Farabaugh 1996; Giedroc et al. 2000; Brierley et al. 2007; Bekaert et al. 2010). In a recent genome-wide analysis on all of the known or expected frameshift sites in animal viruses, we found that  $\sim 85\%$  of the sites have potential H-type pseudoknots downstream from the slippery sequences. Some of the pseudoknots we detected were not known previously. Of particular interest, several potential pseudoknots were detected in the *gag-pol* frameshift junction of the reference strain of HIV-1 (strain HXB2, GenBank accession No. K03455), suggesting possible involvement of pseudoknots in  $-1$  PRF of HIV-1.

In this study, we have analyzed 4254 sequences at the *gag-pol* frameshift junction of all HIV-1 strains whose full-length genome sequences are available in the HIV databases (<http://www.hiv.lanl.gov/>). Our results reveal that the *gag-pol* junction sequences of HIV-1 have the potential to harbor several highly conserved pseudoknots. We propose a novel mechanism of  $-1$  PRF in HIV-1 to explain the possible involvement of these putative pseudoknots.

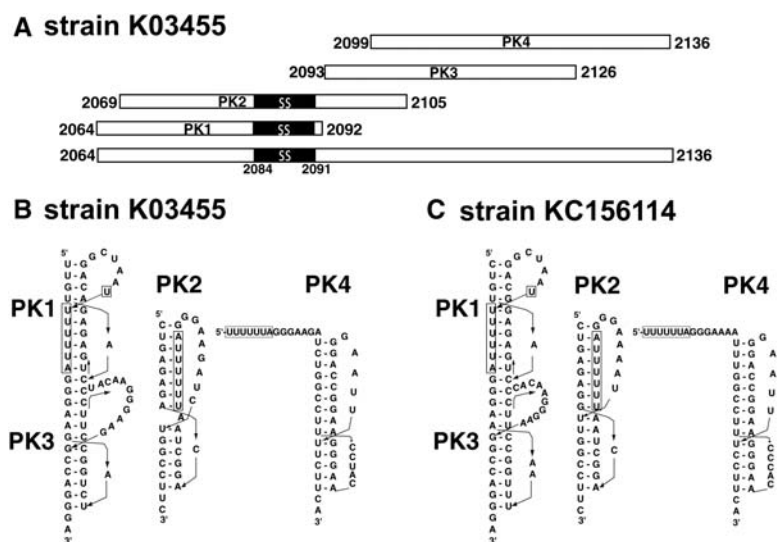
## RESULTS

To study the potential frameshifting signals in the *gag-pol* junction of HIV-1, a sequence window of 127 nt containing the slippery sequence (UUUUUA) plus 50 nt upstream of and 70 nt downstream from the slippery sequence is used. There are 4272 full-length genomic sequences for different strains of HIV-1 in the HIV sequence database. Excluding 18 sequences that either contain nonstandard letter(s) within the 127-nt *gag-pol* sequence window or do not have the UUUUUA frameshift site, 4254 sequences were analyzed. These sequences from different HIV-1 strains provide an excellent and rigorous data set for bioinformatic studies on RNA structures within the HIV-1 genome (in our case, possible involvement of pseudoknots in  $-1$  PRF).

An in-house-developed program (Huang et al. 2013) was used to detect potential H-type pseudoknots within the *gag-pol* sequence window. Four po-

tential pseudoknots were detected in most ( $\sim 95\%$ ) of the HIV-1 strains. To facilitate description and discussion, we refer to the four potential pseudoknots as PK1, PK2, PK3, and PK4. A schematic diagram of the pseudoknot-forming sequences of the four pseudoknots in the reference strain HXB2 is shown in Figure 1A. The predicted secondary structures of the four pseudoknots are shown in Figure 1, B and C for two representative strains—K03455 and KC156114, respectively.

In general, PK1 encompasses the slippery sequence UUUUUA at the 3' end of the pseudoknot-forming sequence. All but one of the nucleotides of the slippery sequence reside within stem 2 of the pseudoknot. PK2 also encompasses the slippery sequence, but it is located in the middle of the pseudoknot-forming sequence; all nucleotides of the slippery sequence reside within stem 1 of the pseudoknot. PK3 follows the slippery sequence immediately (without a spacer region or a minimal 1- to 2-nt spacer region between the pseudoknot and the slippery sequence). PK4 locates 7 nt downstream from the slippery sequence. PK1 and PK3 are arranged in a tandem manner. The stem regions of the two pseudoknots have the potential to stack and form a quasi-continuous helix with a total of 22 bp. As shown in Figure 1A, the pseudoknot-forming sequence of PK2 overlaps with the pseudoknot-forming sequences of all the other three pseudoknots, i.e., PK2 cannot form simultaneously with any one of PK1, PK3, and PK4 in the same viral RNA sequence. The pseudoknot-forming sequences of PK3 and PK4 also overlap with each other; and thus PK3 and PK4 cannot form simultaneously.



**FIGURE 1.** Four putative pseudoknots (PK1–PK4) at the *gag-pol* frameshift junction of HIV-1. (A) A schematic diagram for the four pseudoknot-forming sequences of PK1–PK4 in the reference strain HXB2 (GenBank accession No. K03455). The slippery sequence (ss, residues 2084–2091) UUUUUA is highlighted in black. The numbers indicate the starting and ending residues of the pseudoknot-forming sequences. (B,C) Predicted secondary structures for the four putative pseudoknots PK1–PK4 in two representative strains. The slippery sequence is boxed.

HIV-1 is notorious for its high mutation rate. The sequences at the *gag-pol* junction from different strains of HIV-1 also exhibit a fair degree of sequence variation, despite the functional importance of the junction. Interestingly, the four potential pseudoknots are highly conserved among the 4254 HIV-1 strains. Approximately ninety-five percent of the strains have the potential to form all four pseudoknots, although the possible secondary structures of the pseudoknots may differ in different strains (see Fig. 1B,C for examples; a complete record of detected pseudoknots in the different strains of HIV-1 can be found in the Supplemental Material). In ~97% of the strains, PK1, PK3, and PK4 are detected. Our results indicate that most of the HIV-1 strains have the potential to form a set of highly conserved pseudoknots at the *gag-pol* frameshift junction.

It was noticed previously that many naturally occurring pseudoknots belonged to a structurally related pseudoknot family known as CPK-1, standing for common pseudoknot motif 1 (Du et al. 1996; Du and Hoffman 1997). A typical CPK-1 pseudoknot has a stem 2 of 6 or 7 bp and a very short loop 1 of 1–2 nt; there is no intervening sequence between the two helical stems; therefore, they can stack to form a quasi-continuous helix. Due to the natural twist of an A-form RNA helix, the distance across the major groove reaches a minimum when the helical stem has 6 or 7 bp. This distance can be bridged by a minimal number of 1–2 nt, with the base(s) of the nucleotide(s) being embedded inside the major groove. Probably for these reasons, CPK-1 is favored in nature as a compact and stable pseudoknotted fold of RNAs. Interestingly, the potential pseudoknots PK1, PK2, and PK3 in most HIV-1 strains conform to the CPK-1 family (Fig. 1B,C; Supplemental Material). The predicted secondary structures of these pseudoknots are comparable to those of the autoregulatory pseudoknot within the gene 32 mRNA of bacteriophages T2/T6 (Du et al. 1996; Du and Hoffman 1997) and the –1 PRF stimulator pseudoknot at the *gag-pro* frameshift junction of simian retrovirus-1 (SRV-1) (Du et al. 1997; Michiels et al. 2001), two of the prototypic members of the CPK-1 family whose three-dimensional structures have been determined.

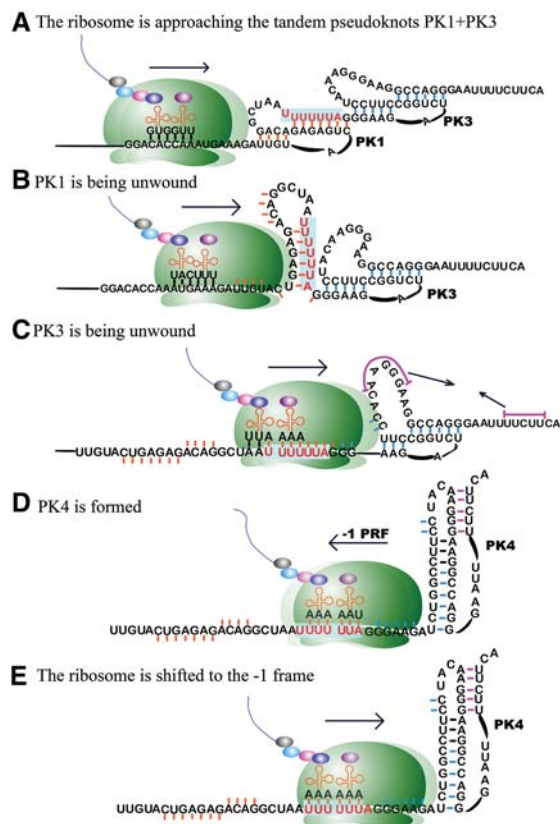
In a typical case of –1 PRF, the stimulator RNA structure locates 5–9 nt downstream from the slippery sequence. With this optimal length of the spacer region, the stimulator RNA structure is positioned at the entrance of the mRNA tunnel of the ribosome, while the slippery sequence is at the A-site and P-site within the ribosome (Namy et al. 2006). The spatial relationship between the slippery sequence and PK4 in most of the HIV-1 strains is typical for –1 PRF. The possible existence of the other three pseudoknots PK1, PK2, and PK3 is unusual. There is no established case for the regulation of –1 PRF by pseudoknots that encompass the slippery sequence or immediately follow the slippery sequence without a spacer region.

The high degree of conservation of the four pseudoknots at the *gag-pol* junction among different HIV-1 strains suggests

that these pseudoknots may participate in the regulation of –1 PRF efficiency of the virus. What are the possible mechanisms for the involvement of these pseudoknots? First, although the detailed mechanisms of –1 PRF have not been fully elucidated, it is generally accepted that the two frameshifting signals, namely the slippery sequence and the downstream stimulator structure, should be separated by a spacer region with 5–9 nt (Baranov et al. 2002; Namy et al. 2006; Brierley et al. 2007). The core event of –1 PRF in HIV-1 should not deviate drastically from this paradigm. Second, due to overlapping of the pseudoknot-forming sequences, certain pairs of pseudoknots (PK1–PK2, PK2–PK3, PK2–PK4, and PK3–PK4) cannot coexist in the same viral mRNA. Keeping these two basic considerations in mind, we propose the following mechanism to explain the plausible involvement of the putative pseudoknots in –1 PRF of HIV-1.

The *gag-pol* frameshift junction of HIV-1 may be structurally heterogeneous, with several alternative structures in equilibrium. In most strains, the PK1 and PK3 tandem pseudoknots may dominate the RNA pool due to their greater overall stability. For example, in the reference strain HXB2, the calculated free energy ( $\Delta G_{37}^{\circ}$ ) value for the stem regions of the tandem pseudoknots is –38.6 kcal/mol, while this value is –20.5 and –26.7 kcal/mol for PK2 and PK4, respectively. PK3 and PK4 overlap in such a way that the stem 2 of PK3 and a large part of the stem 1 of PK4 are the same. The three pseudoknots PK1, PK3, and PK4 may work synergistically to regulate –1 PRF. It is known that some mRNA structures, especially pseudoknots, can cause the translating ribosome to pause upstream of such structures (Tu et al. 1992; Somogyi et al. 1993; Kontos et al. 2001). Assuming that the tandem pseudoknots PK1 and PK3 are present at the *gag-pol* frameshift junction, these pseudoknots may signal the translating ribosome to slow down (Fig. 2A). Since pseudoknots are harder to unwind by the translating ribosomes than simple stem-loop structures with comparable thermodynamic stability (Plant and Dinman 2005), PK1 and PK3 may function as a braking system to slow down the ribosome. To access the slippery sequence which is located at the 3' end of PK1, the translating ribosome has to first unwind PK1 (Fig. 2B). When the ribosome reaches the slippery sequence, PK3 begins to be unwound (Fig. 2C). When the slippery sequence occupies the A- and P-sites, the two 0 frame codons UUU and UUA are recognized by the A- and P-site tRNAs. Stem 1 of PK3 is unwound and the 5' end sequence of PK3 enters the mRNA tunnel of the ribosome. Stem 2 of PK3 remains intact. The 6 bp of PK3 stem 2 are actually the same as the first 6 bp of PK4 stem 1. PK3 and PK4 are mutually excluded due to sequence overlap. Now that PK3 stem 1 is disrupted, PK4 can form rapidly (because a large portion of its stem 1 is already in place). The newly formed PK4, with an optimal spacer from the slippery sequence, jams the entrance of the mRNA tunnel of the translating ribosome (Fig. 2D). At this point, the system reaches the frameshift-





**FIGURE 2.** A proposed mechanism (A–E) for the involvement of the putative pseudoknots PK1, PK3, and PK4 in the regulation of  $-1$  PRF in HIV-1. The sequence of strain HXB2 is used as an example. The secondary structures of the pseudoknots are the same as in Figure 1, although the drawings appear different. The slippery sequence UUUUUUA is highlighted in red. Small bars associated with the residues indicate that the residues are or were participating in base-pairing interactions.

ready stage and proceeds to induce  $-1$  PRF as in a typical case (Fig. 2E).

If PK2 is present in a particular viral mRNA sequence, all the other three pseudoknots would not form in the same sequence at the same time. The process of unwinding PK2 may also significantly slow down the translating ribosome. After PK2 is disrupted, PK4 can form. Due to the peculiar location of the slippery sequence within PK2 and the way in which the pseudoknot-forming sequences of PK2 and PK4 are overlapped (Fig. 1), formation of PK4 may happen before the translating ribosome reaches the slippery sequence. In this scenario, the role of PK2 is similar to that of the PK1–PK3 tandem pseudoknots, albeit with a lesser degree of complexity and elaboration.

In the above scenarios, PK4 is not present originally. It forms only after the disruption of PK2 or PK3 by the translating ribosome. However, it is also possible that PK4 is present in the first place. In this case, PK2 and PK3 cannot form in the same mRNA sequence at the same time. PK1 can coexist with PK4; these two pseudoknots are separated

by a short linker region with 6 nt (in strains K03455 and KC156114) (Fig. 1B,C). The translating ribosome has to unwind PK1 and may slow down in the unwinding process.

Due to the overlapping nature of the putative pseudoknots at the *gag-pol* junction of HIV-1 mRNAs, several scenarios are possible for the involvement of these pseudoknots. The relative stabilities of the pseudoknots may decide which scenario dominates. Other *trans*-acting factors may also participate in the regulatory processes by stabilizing a particular pseudoknot or set of pseudoknots. Involvement of a cellular factor (eukaryotic release factor 1) in the modulation of HIV-1 PRF has been reported (Kobayashi et al. 2010).

The frameshift-stimulating secondary structure downstream from the HIV-1 group M (which includes the reference strain HXB2) *gag-pol* frameshift site was originally proposed to be a simple stem-loop (Jacks et al. 1988), which was shown to be important for wild-type level frameshifting in vivo (in mammalian cells) (Parkin et al. 1992). It was shown later that a sequence downstream from the originally proposed stem-loop also contributed to frameshifting, either modeled as an elaborated pseudoknot with an intra-molecular triplex formed between loop 1 and stem 2 of the pseudoknot (Dinman et al. 2002) or an extended bulged stem-loop (Dulude et al. 2002). The elaborated pseudoknot has the same pseudoknot-forming sequence as PK4 shown in Figure 1. The difference is in the base-pairing schemes and the proposed triple-helix. As pointed out previously (Dulude et al. 2002), the elaborated pseudoknot with a triple-helix is highly improbable because it is sterically disfavored. The extended bulged stem-loop structure covers the sequence that forms PK4, plus the purine-rich spacer region. In the structure, the first 3 nt (GAA) in loop 2 of PK4 form a bulge; the pyrimidine-rich sequence 3' to the bulge forms a stem with the purine-rich spacer region instead of forming stem 2 of PK4, as shown in Figure 1, B and C. In the extended bulged stem-loop model, there is no spacer region between the slippery sequence and the downstream RNA structure. An A-C mismatch is present in the stem in 38% of the 139 sequences analyzed (Dulude et al. 2002).

In HIV-1 strain MVP5180 (GenBank accession No. L20571) from subgroup O, a very classic H-type pseudoknot (with 8 bp in both stems) locating 8 nt downstream from the *gag-pol* frameshift site was shown to be required for stimulating a higher frameshifting efficiency than that in group M (Baril et al. 2003). This established pseudoknot corresponds to PK4 in the strains shown in Figure 1 (see the Supplemental Material). Interestingly, the PK4 in strain MVP5180 (calculated  $\Delta G_{37}^{\circ}$  for the stems =  $-40.1$  kcal/mol) is expected to be much more stable than the PK4 in strains HXB2 (calculated  $\Delta G_{37}^{\circ}$  for the stems =  $-26.7$  kcal/mol). There are 28 strains of subgroup O in this analysis. The PK4 in strain MVP5180 is conserved in most subgroup O strains. PK2 is not detected in subgroup O strains. About half of the subgroup O strains have PK1 and PK4; the other half of the strains also have PK3 (see the Supplemental Material for

strains L20571, L20587, JX245014-15, JN571034, AY623602, AY618998, AY169802-16, AJ302646-47, AF407418, AB485666-69). Due to its greater stability, PK4 may dominate the RNA pool over other alternative structures in subgroup O. The essence here is that our proposed mechanism incorporates strains in subgroup O within the same framework as all the other strains of HIV-1.

Compared to the previously proposed triplex-containing pseudoknot and extended bulged stem-loop models, the three or four putative pseudoknots identified in this study have a higher degree of conservation. This high degree of conservation strongly suggests that these pseudoknots may assume functional roles.

## DISCUSSION

Results from our analysis and previous studies indicate that the RNA sequence at the *gag-pol* frameshift junction of HIV-1 has the potential to form a number of alternative structures, including the pseudoknots presented in this study and the previously characterized triplex-containing pseudoknot and stem-loop (Dinman et al. 2002; Dulude et al. 2002). Some of these structures cannot coexist in the same sequence. These observations raise the intriguing possibility that the HIV-1 *gag-pol* junction sequence may be structurally heterogeneous. An equilibrium may exist for multiple alternative structures. Depending on the particular RNA structures present, a specific mRNA sequence may or may not be active in -1 PRF. In a previous study on the murine leukemia virus read-through stimulating pseudoknot (MLV-PK), it was shown that MLV-PK is equilibrated between an active (PK<sub>active</sub>) and an inactive (PK<sub>inactive</sub>) conformation. At physiological pH, the read-through permissive PK<sub>active</sub> conformation accounted for 6% of the population (Houck-Loomis et al. 2011). A somewhat similar equilibrium-based mechanism may play a role in -1 PRF of HIV-1. In the case of the murine leukemia virus, two different conformations of the same pseudoknot structure are in equilibrium, while in the putative case of HIV-1, different RNA structures are involved, which may further have different conformations.

It is reported that the 5' UTR of HIV-1 genomic RNA can exist in two alternative structures: the branched multiple-hairpin (BMH) structure and the long distance interaction (LDI) structure (Ooms et al. 2004; Abbink et al. 2005). Switching between the BMH and LDI conformations requires very significant structural rearrangements. The mechanisms regulating this riboswitch are not well understood. The BMH-LDI riboswitch is not related to the putative alternative pseudoknot structures at the *gag-pol* junction due to different locations in the viral genomic RNA. However, existence of the BMH-LDI riboswitch indicates that functionally important regions in the HIV-1 genomic RNA can adopt multiple interconverting structures.

Our results also suggest the possible involvement of sequence upstream of the slippery sequence in the regulation

of -1 PRF in HIV-1. Although most previous studies on -1 PRF signals have focused on sequence downstream from the slippery sequence, a few studies have revealed that upstream sequences can also affect -1 frameshifting efficiency. In HIV-1 and HTLV-2 (Human T-cell leukemia virus-2), the 8 nt immediately upstream of the slippery sequences were shown to enhance or attenuate frameshifting in a context-dependent manner (Kim et al. 2001). It was also found that the E-site codon preceding the slippery sequence in HIV-1 was involved in the regulation of -1 PRF (Leger et al. 2007). In barley yellow dwarf virus (BYDV), a conserved stem-loop structure immediately upstream of the slippery sequence enhanced frameshift efficiency. The proposed role of the upstream structure is to slow down the translating ribosome (Barry and Miller 2002). In SARS CoV, frameshifting activity is down-regulated by a stem-loop immediately upstream of the slippery sequence (Cho et al. 2013). In a SHAPE-derived secondary structure model of the entire HIV-1 genomic RNA (Watts et al. 2009), the *gag-pol* frameshift junction was modeled as a three-helix junction structure. Interestingly, the slippery sequence pairs with an upstream sequence to form one of the three helices. It should be noted that SHAPE-based RNA secondary structure prediction has several limitations (Low and Weeks 2010; Kladwang et al. 2011), most of which are highly relevant to comparing the SHAPE-derived structure and the pseudoknot structures proposed in this manuscript. First, the dynamic programming algorithms used for SHAPE-based secondary structure prediction exclude pseudoknots. Second, SHAPE data only reflect the bulk behavior of the RNA sample; therefore, SHAPE-based secondary structure prediction is inadequate for characterizing RNA structures in interconverting states. Third, the biologically active structure may not always be the one that dominates the RNA pool (with minimum free energy).

No report has suggested the involvement of upstream sequence in the form of a pseudoknot. The detected pseudoknots in our study not only encompass the upstream sequence but also the slippery sequence and sequence corresponding to the spacer region in a typical case. Moreover, in our proposed mechanism, multiple pseudoknots function synergistically in the process. In the presumably dominating scenario (in most strains) as shown in Figure 2, PK1 and PK3 are arranged tandemly. Such tandem structures have not been implicated in -1 PRF and other decoding processes. PK3 and PK4 are mutually excluded. Unwinding of PK3 enables PK4 to form. The process is, therefore, dynamic, in which the translating ribosome plays an active role in unwinding the PK1/PK3 tandem pseudoknots and inducing the formation of PK4 that is the end effector structure for frameshifting stimulation. This novel mechanism of pseudoknot-mediated -1 PRF in HIV-1 has many exquisite features that have not been described in any of the previously characterized cases of -1 PRF, but the mechanism is still consistent with the current paradigm of pseudoknot-dependent -1 PRF.

## MATERIALS AND METHODS

The HIV-1 sequences were downloaded from the HIV databases (<http://www.hiv.lanl.gov/>). There are 4272 full-length genomic sequences in the April 18, 2013 download.

All HIV-1 strains harbor a slippery sequence UUUUUUA at the *gag-pol* frameshift junction. For our analysis, a sequence window containing 50 nt upstream of the slippery sequence and 70 nt downstream from the slippery sequence is used. Potential H-type pseudoknots within the sequence window are detected by an in-house-developed program. Details of the program have been described elsewhere (Huang et al. 2013). Briefly, a typical H-type pseudoknot contains two double-stranded regions (stem 1 and stem 2) separated by two or three single-stranded connecting sequences (loop 1, loop 2, and optionally, loop 3). Within user-defined ranges of stem and loop lengths, the computer program tests all possible combinations of stem and loop lengths to see whether the given RNA sequence harbors all the necessary sequence elements for pseudoknot formation. In this study, the ranges of stem and loop lengths for pseudoknot detection are as follows: stem 1 and stem 2: 3–20 bp; loop 1: 1–15 nt; loop 2: 3–50 nt; loop 3: 0–10 nt. No mismatched pair (G-U is considered a legitimate base pair) or bulge is allowed within the stems. Therefore, the detected pseudoknots all have perfectly complementary stems. To get a rough idea about the relative thermodynamic stability of the detected pseudoknots, free energy ( $\Delta G_{37}^{\circ}$ ) values for the two stems S1 and S2 are calculated. Turner's nearest-neighbor parameters are used in the calculation (Serra and Turner 1995). If loop 3 is absent, the two stems are treated as a continuous helical stem, but only half of the value is given to the stem 1–stem 2 stack to account for the quasi-continuous nature of the stacked stems.

## SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

## ACKNOWLEDGMENTS

This work was supported by the start-up fund and a seed grant from Southern Illinois University Carbondale to Z.D. and in part by the National Science Foundation under Grant No. IIS-1218712 to Q.C. Support also came from a Dissertation Research Award to X.H. from the graduate school of Southern Illinois University Carbondale and a Gower award to G.W. from the Department of Chemistry and Biochemistry of Southern Illinois University Carbondale.

Received September 11, 2013; accepted January 25, 2014.

## REFERENCES

- Abbinck TE, Ooms M, Haasnoot PC, Berkhout B. 2005. The HIV-1 leader RNA conformational switch regulates RNA dimerization but does not regulate mRNA translation. *Biochemistry* **44**: 9058–9066.
- Atkins JF, Weiss RB, Gesteland RF. 1990. Ribosome gymnastics—degree of difficulty 9.5, style 10.0. *Cell* **62**: 413–423.
- Baranov PV, Gesteland RF, Atkins JF. 2002. Recoding: translational bifurcations in gene expression. *Gene* **286**: 187–201.
- Baril M, Dulude D, Steinberg SV, Brakier-Gingras L. 2003. The frameshift stimulatory signal of human immunodeficiency virus type 1 group O is a pseudoknot. *J Mol Biol* **331**: 571–583.
- Barry JK, Miller WA. 2002. A –1 ribosomal frameshift element that requires base pairing across four kilobases suggests a mechanism of regulating ribosome and replicase traffic on a viral RNA. *Proc Natl Acad Sci* **99**: 11133–11138.
- Bekaert M, Firth AE, Zhang Y, Gladyshev VN, Atkins JF, Baranov PV. 2010. Recode-2: new design, new search tools, and many more genes. *Nucleic Acids Res* **38**: D69–D74.
- 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.
- Brierley I. 1995. Ribosomal frameshifting viral RNAs. *J Gen Virol* **76**: 1885–1892.
- Brierley I, Pennell S, Gilbert RJ. 2007. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Microbiol* **5**: 598–610.
- Cho CP, Lin SC, Chou MY, Hsu HT, Chang KY. 2013. Regulation of programmed ribosomal frameshifting by co-translational refolding RNA hairpins. *PLoS One* **8**: e62283.
- Dam E, Pleij K, Draper D. 1992. Structural and functional aspects of RNA pseudoknots. *Biochemistry* **31**: 11665–11676.
- Dinman JD, Richter S, Plant EP, Taylor RC, Hammell AB, Rana TM. 2002. The frameshift signal of HIV-1 involves a potential intramolecular triplex RNA structure. *Proc Natl Acad Sci* **99**: 5331–5336.
- Du Z, Hoffman D. 1997. An NMR and mutational study of the pseudoknot within the gene 32 mRNA of bacteriophage T2: insights into a family of structurally related RNA pseudoknots. *Nucleic Acids Res* **25**: 1130–1135.
- Du Z, Giedroc DP, Hoffman DW. 1996. Structure of the autoregulatory pseudoknot within the gene 32 messenger RNA of bacteriophages T2 and T6: a model for a possible family of structurally related RNA pseudoknots. *Biochemistry* **35**: 4187–4198.
- Du Z, Holland JA, Hansen MR, Giedroc DP, Hoffman DW. 1997. Base-pairings within the RNA pseudoknot associated with the simian retrovirus-1 *gag-pro* frameshift site. *J Mol Biol* **270**: 464–470.
- Dulude D, Baril M, Brakier-Gingras L. 2002. Characterization of the frameshift stimulatory signal controlling a programmed –1 ribosomal frameshift in the human immunodeficiency virus type 1. *Nucleic Acids Res* **30**: 5094–5102.
- Farabaugh PJ. 1996. Programmed translational frameshifting. *Microbiol Rev* **60**: 103–134.
- Gareiss PC, Miller BL. 2009. Ribosomal frameshifting: an emerging drug target for HIV. *Curr Opin Investig Drugs* **10**: 121–128.
- Gesteland RF, Atkins JF. 1996. Recoding: dynamic reprogramming of translation. *Annu Rev Biochem* **65**: 741–768.
- Gesteland RF, Weiss RB, Atkins JF. 1992. Recoding: reprogrammed genetic decoding. *Science* **257**: 1640–1641.
- 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.
- Houck-Loomis B, Durney MA, Salguero C, Shankar N, Nagle JM, Goff SP, D'Souza VM. 2011. An equilibrium-dependent retroviral mRNA switch regulates translational recoding. *Nature* **480**: 561–564.
- Huang X, Du Z, Cheng J, Cheng Q. 2013. PKscan: a program to identify H-type RNA pseudoknots in any RNA sequence with unlimited length. *Bioinformatics* **9**: 440–442.
- Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE. 1988. Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* **331**: 280–283.
- 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.
- Kladwang W, VanLang CC, Cordero P, Das R. 2011. Understanding the errors of SHAPE-directed RNA structure modeling. *Biochemistry* **50**: 8049–8056.
- Kobayashi Y, Zhuang J, Peltz S, Dougherty J. 2010. Identification of a cellular factor that modulates HIV-1 programmed ribosomal frameshifting. *J Biol Chem* **285**: 19776–19784.

- Kontos H, Naphine 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.
- Leger M, Dulude D, Steinberg SV, Brakier-Gingras L. 2007. The three transfer RNAs occupying the A, P and E sites on the ribosome are involved in viral programmed -1 ribosomal frameshift. *Nucleic Acids Res* **35**: 5581–5592.
- Low JT, Weeks KM. 2010. SHAPE-directed RNA secondary structure prediction. *Methods* **52**: 150–158.
- 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.
- Marcheschi RJ, Tonelli M, Kumar A, Butcher SE. 2011. Structure of the HIV-1 frameshift site RNA bound to a small molecule inhibitor of viral replication. *ACS Chem Biol* **6**: 857–864.
- McNaughton BR, Gareiss PC, Miller BL. 2007. Identification of a selective small-molecule ligand for HIV-1 frameshift-inducing stem-loop RNA from an 11,325 member resin bound dynamic combinatorial library. *J Am Chem Soc* **129**: 11306–11307.
- 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.
- Namy O, Moran SJ, Stuart DI, Gilbert RJC, Brierley I. 2006. A mechanical explanation of RNA pseudoknot function in programmed ribosomal frameshifting. *Nature* **441**: 244–247.
- Ooms M, Huthoff H, Russell R, Liang C, Berkhout B. 2004. A riboswitch regulates RNA dimerization and packaging in human immunodeficiency virus type 1 virions. *J Virol* **78**: 10814–10819.
- Park SJ, Kim YG, Park HJ. 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.
- Parkin NT, Chamorro M, Varmus HE. 1992. Human immunodeficiency virus type 1 *gag-pol* frameshifting is dependent on downstream mRNA secondary structure: demonstration by expression in vivo. *J Virol* **66**: 5147–5151.
- Plant EP, Dinman JD. 2005. Torsional restraint: a new twist on frameshifting pseudoknots. *Nucleic Acids Res* **33**: 1825–1833.
- Pleij CW. 1990. Pseudoknots: a new motif in the RNA game. *Trends Biochem Sci* **15**: 143–147.
- Pleij CW, Rietveld K, Bosch L. 1985. A new principle of RNA folding based on pseudoknotting. *Nucleic Acids Res* **13**: 1717–1731.
- Serra MJ, Turner DH. 1995. Predicting thermodynamic properties of RNA. *Methods Enzymol* **259**: 242–261.
- Somogyi P, Jenner AJ, Brierley I, Inglis SC. 1993. Ribosomal pausing during translation of an RNA pseudoknot. *Mol Cell Biol* **13**: 6931–6940.
- Tu C, Tzeng TH, Bruenn JA. 1992. Ribosomal movement impeded at a pseudoknot required for frameshifting. *Proc Natl Acad Sci* **89**: 8636–8640.
- Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW Jr, Swanstrom R, Burch CL, Weeks KM. 2009. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature* **460**: 711–716.