



Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.

## ORIGINAL ARTICLES

## RNA Pseudoknots Downstream of the Frameshift Sites of Retroviruses

SHU-YUN LE, BRUCE A. SHAPIRO, JIH-H. CHEN, RUTH NUSSINOV, and JACOB V. MAIZEL

*RNA pseudoknot structural motifs could have implications for a wide range of biological processes of RNAs. In this study, the potential RNA pseudoknots just downstream from the known and suspected retroviral frameshift sites were predicted in the Rous sarcoma virus, primate immunodeficiency viruses (HIV-1, HIV-2, and SIV), equine infectious anemia virus, visna virus, bovine leukemia virus, human T-cell leukemia virus (types I and II), mouse mammary tumor virus, Mason-Pfizer monkey virus, and simian SRV-1 type-D retrovirus. Also, the putative RNA pseudoknots were detected in the gag-pol overlaps of two retrotransposons of Drosophila, 17.6 and gypsy, and the mouse intracisternal A particle. For each sequence, the thermodynamic stability and statistical significance of the secondary structure involved in the predicted tertiary structure were assessed and compared. Our results show that the stem-loop structures in the pseudoknots are both thermodynamically highly stable and statistically significant relative to other such configurations that potentially occur in the gag-pol or gag-pro and pro-pol junction domains of these viruses (300 nucleotides upstream and downstream from the possible frameshift sites are included). Moreover, the structural features of the predicted pseudoknots following the frameshift site of pro-pol overlaps of the HTLV-1 and HTLV-2 retroviruses are structurally well conserved. The occurrence of eight compensatory base changes in the tertiary interaction of the two related sequences allow the conservation of their tertiary structures in spite of the sequence divergence. The results support the possible control*

From the Institute of Biological Sciences (S.-Y.L.), National Research Council of Canada, Ottawa, Canada; and the Laboratory of Mathematical Biology (B.A.S., R.N., J.V.M.), Division of Cancer Biology, Diagnosis, and Centers, National Cancer Institute, National Institutes of Health, Frederick, Maryland, USA; and the Advanced Scientific Computer Laboratory (J.-H.C.), Program Resources, Inc., National Cancer Institute/FCRDC, Frederick, Maryland, U.S.A.

Address correspondence to Dr. J. V. Maizel, Chief, Laboratory of Mathematical Biology, Division of Cancer Biology, Diagnosis, and Centers, National Cancer Institute, PO Box B, Room 151, Building 469, Frederick, MD 21702-1201, USA.

Received 15 April 1991; revised and accepted 30 July 1991.

*mechanism for frameshifting proposed by Brierley et al. [1] and Jacks et al. [2, 3].*

## Introduction

It has been demonstrated that pseudoknots are significant structural motifs that are found in ribosomal RNAs [4], mRNA leader sequences [5], and viral RNAs [6–11], especially in plant viral tRNA-like structures [12]. Although the precise three-dimensional conformation of the pseudoknot remains uncertain, it can be considered to form a long quasi-continuous coaxial stacking region [12, 13]. This possible coaxial stacking region was suggested to be a common feature in the structure of a number of plant viral RNA 3' termini [14]. Pseudoknots have been implicated in several biological phenomena involving the replication of the plus strand of plant viruses [14], interaction between proteins and RNA [15, 16], and efficient ribosomal frameshifting [1]. Among them, the site-specific mutagenesis approach within the frameshift region of the genomic RNA of the avian coronavirus infectious bronchitis virus (IBV) has revealed that the RNA pseudoknot just downstream of the frameshift site is an essential element of the IBV ribosomal frameshifting signal [1].

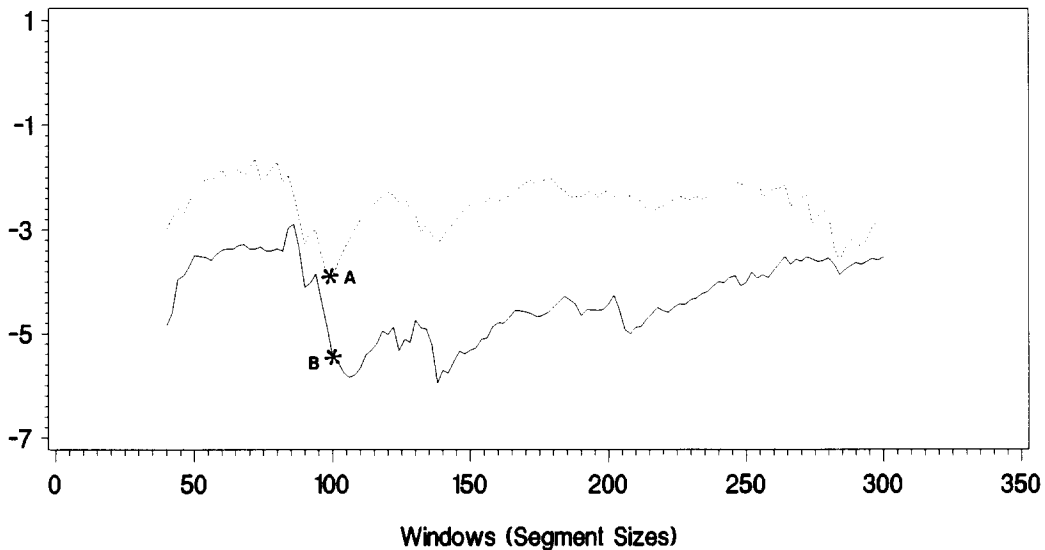
Retroviral ribosomal frameshifting involving ribosomal slippage into a  $-1$  reading frame in gag-pol (or gag-pro-pol) overlapping domains is commonly used in the native retroviral translational system. This translation mechanism clearly benefits retroviruses in that one kind of mRNA molecule can direct large amounts of structural (gag) protein synthesis and relatively small amounts of catalytic (pro and pol) protein synthesis, while the attached gag components can direct incorporation of pol products into viral cores. The basis of such translation control has been proposed [2, 3, 17–19]. Current studies have demonstrated that a “slippery” sequence at the ribosomal frameshift site and a stem-loop structure just downstream from it are essential for the frameshifting of several retroviruses. More recently, we have analyzed RNA secondary structures [20] that are predicted in the gag-pol or gag-pro and pro-pol junction regions (600 nucleotides) of Rous sarcoma virus (RSV), human immunodeficiency virus (HIV-1), bovine leukemia virus (BLV), human T-cell leukemia virus type II (HTLV-2), and mouse mammary tumor virus (MMTV) by a recently developed Monte Carlo simulation method [21, 22]. Extensive simulations revealed that pre-

dicted stem-loops just downstream of the “slippery” sequence are thermodynamically highly stable and their involvement in frameshifting is statistically relevant relative to other such configurations that may occur in these junction regions.

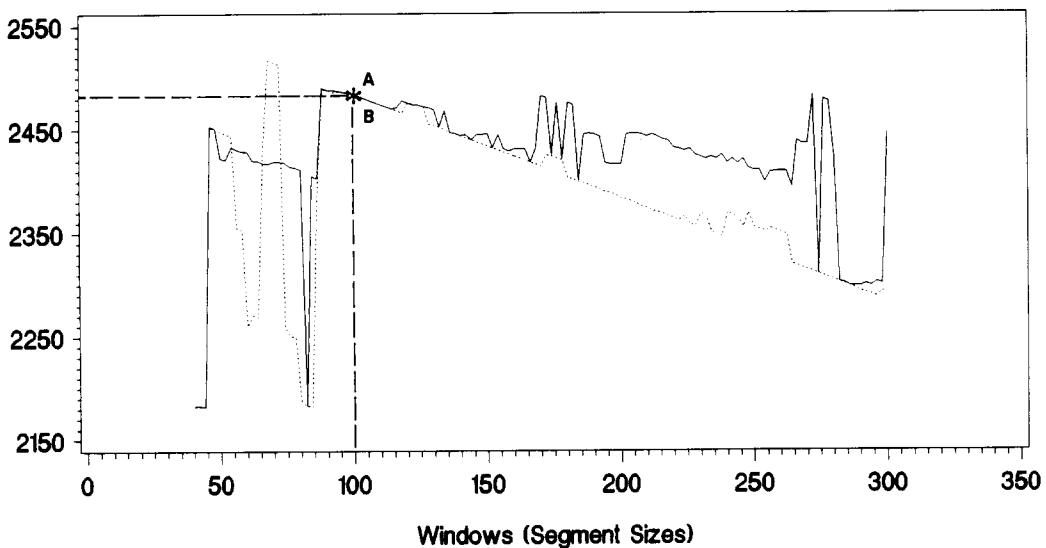
Although a complex structure termed a pseudoknot has been demonstrated in the efficient ribosomal frameshifting in the nonretroviral system of IBV [1], it is unclear whether the pseudoknot structural motif is a general signal for the ribosomal frameshifting event of the retroviral

system. In a recent report, Jacks et al. [3] conjectured that such a complex pseudoknot structure may occur immediately 3' to the frameshift site of RSV. This structure comprises tertiary interactions between single-stranded nucleotides in the stem-loop and nucleotides downstream of this stem-loop structure. Furthermore, Brierley et al. [1] have also suspected that the tertiary RNA interactions may be a common feature of (-1) frameshift sites by searching for potential pseudoknots in retroviruses. However, the basis of these predicted pseudoknot structures is not

**a** Minimum Scores



**b** Starting Positions



clear. The thermodynamic stability and statistical relevance of the structural features in the predicted pseudoknots relative to other structures that potentially occur within the upstream or downstream of the frameshift site are not assessed. The details about the structural features of these pseudoknots are also not fully known.

In this study, we analyze all 13 different species of retroviruses and three transposable elements in which the known or suspected frameshift sites have been proposed by Jacks et al. [3]. These are RSV [23], HIV-1 [24], HIV-2 [24], simian immunodeficiency virus (SIVmac and SIVagm) [24], equine infectious anemia virus (EIAV) [24], visna virus (VISNA) [24], BLV [25], HTLV-1 [24], HTLV-2 [24], MMTV [2, 17], Mason-Pfizer monkey virus (MPMV) [26], simian SRV-1 type-D retrovirus (SRV-1) [27], and two retrotransposons of *Drosophila*, 17.6 [28] and gypsy [29], and the mouse intracisternal A particle (MIAP) [30]. Among them, different isolate sequences [24] of HIV-1, HIV-2, and SIVmac are also included. Putative RNA pseudoknots just downstream of the demonstrated and suspected frameshift sites are detected in these sequences by an extensive computer simulation. In the simulation analysis, the extended overlapping domain or junction domain (that is, gag-pol, or gag-pro and pro-pol domains) is selected to include 300 nucleotides (nt) both upstream and downstream from the potential frameshift site. The thermodynamic stability and statistical significance of all possible secondary structures that potentially occur in the extended

overlapping domain are examined. The stem-loop structures entangled in the predicted RNA pseudoknots are both highly stable and statistically relevant relative to other structures within these junction domains. The structural features of the predicted pseudoknots following the frameshift site of pro-pol overlaps of the HTLV-1 and HTLV-2 retroviruses are structurally well conserved. The occurrence of eight compensatory base changes in the tertiary interaction of the two related sequences allow the conservation of their tertiary structures in spite of the sequence divergence. The strong structural conservation combined with the structural feature of highly stable and statistically more significant stem-loop involved in the conserved putative pseudoknot suggests that the distinct RNA pseudoknot could be essential for the efficient frameshifting of HTLV-1 and HTLV-2 retroviruses.


Similar structures have been recently obtained using a different theoretical approach strengthening the results presented here [31].

## Materials and Methods

In our study, the gag-pol domains of RSV, HIV-1, HIV-2, SIVmac, SIVagm, EIAV, VISNA, 17.6, gypsy, MIAP, and gag-pro and pro-pol domains of BLV, HTLV-1, HTLV-2, MMTV, MPMV, and SRV-1 are determined by their open reading frames and the demonstrated or suspected frameshift sites [3] in the genomes of these retroviruses or transposable elements. These junction regions consist of ~600 nt (that is, they include 300 nt both upstream and downstream from the possible frameshift sites).

The putative pseudoknot is detected in two steps. Step 1 involves the detection of a distinct unusual RNA folding region. The RNA secondary structure predicted in the specific folding region is highly stable and statistically significant relative to others in a given domain of the sequence. In step 2, the possible tertiary interactions between unpaired nucleotides in the loop and nucleotides downstream of the detected distinct stem-loop structure are searched.

The stability and statistical significance of a given stem-loop relative to other possible structures in the sequence are assessed by two standardized scores, stability score and significance score [32, 33]. The significance score (Sigscr) and stability score (Stbscr) of a segment are defined as


**Figure 1.** (a) Distributions of minimal stability score (dashed curve) and significance score (solid curve) in the gag-pol junction region (600 nucleotides [nt]) of RSV under different window sizes. For each window, the simulation was carried out by sliding one base along the RNA sequence. In this way the global minima of stability and significance scores were extracted and plotted against the window size. The exhaustive simulation was completed for window sizes ranging from 40 to 300 nt by increasing the window by 2 nt at a time. The deep valley A shows that the global minimal stability score in the extensive simulation is obtained when the window size is taken as 100 nt. (b) Distributions of highly stable RNA folding segment (dashed curve) and significant folding segment (solid curve) under different window sizes. The map was obtained by plotting the starting positions of these unusual RNA folding segments against the window sizes. When the window is taken as 100 nt, which is suitable to detect a both highly stable and more significant folding region, the starting positions (2484) of the highly stable and of the more significant folding regions coincide with each other, labeled A and B. The corresponding sequence position of the both highly stable and significant region is identified at 2484–2583.

$$\text{Sigschr} = \frac{E - E_r}{SD_r}$$

$$\text{Stbschr} = \frac{E - E_b}{SD_b}$$

In the two equations,  $E$  is the lowest free energy of the real biological sequence in the segment.  $E_r$  and  $SD_r$  are the mean and standard deviation of the lowest free energies from a large number of randomly shuffled sequences of the segment [21, 22].  $E_b$  and  $SD_b$  are the mean and standard deviation of the minimum free energies computed by folding all fragments of the same size within the given sequence [32]. The distributions of the significance and stability scores in the sequence are computed using the program SEG FOLD [32]. The program was vectorized and optimized to run on the Cray operating system of a Cray X-MP/24 supercomputer. Also, the program has been adapted for the VAX/VMS and IRIS/UNIX environments. Currently,  $E_r$  and  $SD_r$  are computed using empirical formulas [34, 35] in all versions of SEG FOLD.

In general, the distinct unusual RNA folding region, which is highly stable and more statistically significant than others in the sequence, is found through an exhaustive Monte Carlo simulation. In this study, the exhaustive simulation of each retrovirus RNA is often performed for window sizes ranging from 40 to 300 nt (in some cases, 30–300 or 20–300 nt is used) by adding 2 nt at a time to the window size. For each window size, the calculation of significance and stability scores is carried out by sliding one base along the RNA sequence in the gag-pol or gag-pro and pro-pol junction domain (600 nt). In the simulation the global minima of the significance score and stability score and their corresponding segment starting positions in the sequence were extracted for each window. By comparison of these minimal scores, the unusual RNA folding region of both high stability and more statistical significance can be detected. For the phylogenetically related retrovirus sequences, when these distinct unusual RNA folding regions are found, the conserved tertiary structure is constructed using a comparative sequence analysis approach.

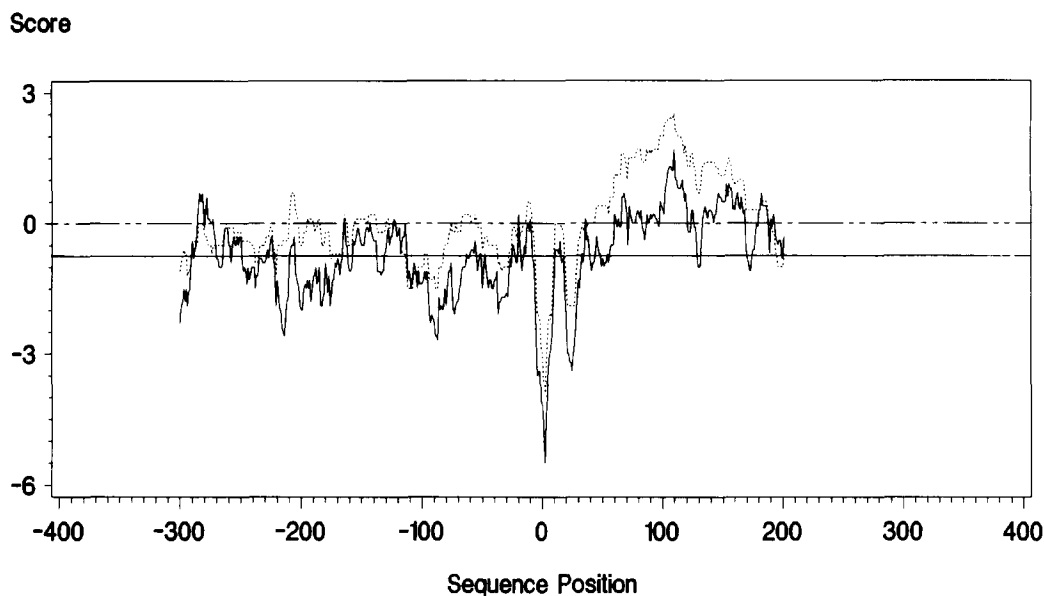
The distinct RNA stem-loop structure in the unusual folding region is predicted using the NEW FOLD program, which is a modified version of Zuker's FOLD program [36]. NEW FOLD incorpo-

rates the Turner energy rules with penalties of asymmetric loops and internal loops closed by AU pairs [37, 38]. The determination of potential RNA pseudoknots was accomplished by a search program developed by Shapiro (in preparation) that currently runs on a SYMBOLICS 3675 Lisp machine. It accepts as input RNA structural information in the form of a region table and its associated sequence and produces as output a listing of all potential pseudoknots formed by loop-free-base interactions. Loop-loop interactions were filtered out in this case since we are dealing with specific significant and highly stable secondary structures formed in the vicinity of the slippage motif. Structural formation beyond this structure is not considered in order to model the paradigm presented by Brierley et al. [1]. Thus, a list of only loop-free-base interactions is presented as a result of the program. This list consists of a quadruple, which is analogous to the quadruples present in the input region table. That is, each quadruple contains the 5' start position, 3' stop position, size, and energy of the potential knot regions. In calculating the latter energy, the unfavorable energy contributions of the pseudoknot loops were not taken into account. However, current practice is to use a constant value, for example, 4.2 kcal/mol [39]. These quadruples are depicted in sorted order increasing with respect to the 3' position so that one may easily judge the distance of the pseudoknot stem from the specified significant RNA secondary structure. In this study, the general scheme used to find potential frameshifting pseudoknots was to take the distinct secondary structures detected in step 1 and append ~100 nt downstream from the structure. The algorithm then generated the sorted putative pseudoknots that formed for each of the sequences. The strongest interactions within a range of ~50 nts downstream taking into account length and energy were used to determine the suggested site.

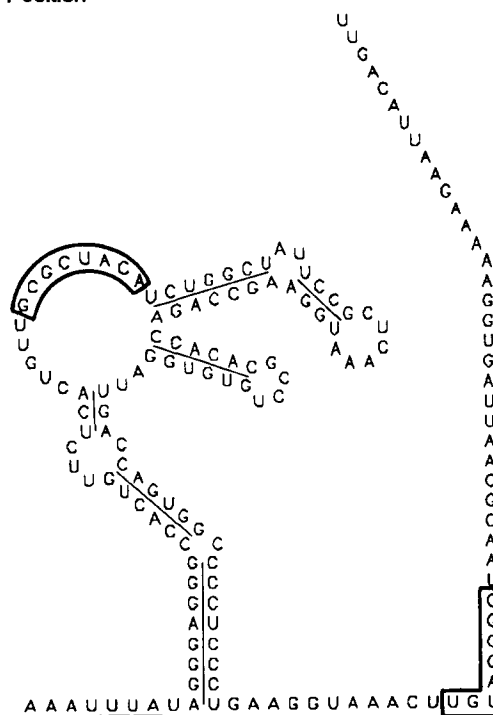
## Results

### *Predicting Highly Stable and Significant Stem-Loops in the gag-pol Junction Domains of Retroviruses*

As an example of our scheme, we present here the highly stable and significant stem-loop detected in the gag-pol domain of RSV. Suitable folding regions for the potential highly significant and stable



**Figure 2.** (a) Distributions of thermodynamic stability (dashed curve) and statistical significance (solid curve) of stem-loop structures in the gag-pol domain of RSV. The stability and significance scores of the stem-loop structures of 100-base-long overlapping segments (window) are plotted against the position of the first base of each window. In the map, the frameshift site [3] is taken as position zero, and horizontal solid and dashed lines represent the average significance and stability score, respectively, in the search region. (b) Potential RNA pseudoknot structure just downstream from the frameshift site of the gag-pol in the RSV. The "slippery" sequence is underlined, and the nucleotides that could participate in tertiary interactions (that is, stem 2 in Table 1) between the loop and downstream region are boxed.



secondary structures within the gag-pol junction domain of RSV were assessed for window sizes ranging from 40 to 300 nt by adding 2 nt to the window size. In the simulation, the global minimum scores of significance and stability and the corresponding starting positions of such highly significant and stable folding regions were computed and sorted for each window size. The relationships between the window size and these global minimum scores (Figure 1a) and between the

window size and these starting positions (Figure 1b) were plotted. In Figure 1a, the deep valley A, where the stability score is  $-3.90$  (global minimum in the distribution plot), can be easily picked up for a window size of 100 nt. We also detected a corresponding deep trough B with the same window size, where the significance score is less than  $-5.0$ . The corresponding highly stable and statistically significant folding regions A and B can be assessed from Figure 1b, while the window size is

taken as 100 nt, determined by the distributions of stability and significance scores (Figure 1a). In this case, the unusual RNA folding regions A and B coincide. Thus, the starting position of the unusual RNA folding region is 2484. The unusual RNA folding region encompassing nucleotides 2484–2583 is thermodynamically highly stable and statistically significant relative to other such configurations that potentially occur within the gag-pol junction domain. A clearer depiction is given in Figure 2a. The distinct RNA secondary structure adjacent to the frameshift site is displayed in Figure 2b. The predicted RNA secondary struc-

ture is similar to the structural model depicted by Jacks et al. [3].

Using the same procedure, all unusual RNA folding regions just downstream of the frameshift sites of other sequences are detected and listed in Table 1 (see the S1 column). The significance and stability scores are listed in columns A and B, respectively. The data presented in Table 1 reveal that the stem-loops predicted downstream from the (–1) ribosomal frameshift sites of gag-pol or gag-pro and pro-pol are also both highly stable and statistically significant in comparison to the others in the junction domain of these retrovirus-

**Table 1.** Putative RNA pseudoknots just downstream from the known or suspected (–1) ribosomal frameshift sites of retroviruses

Virus or transposons	Slippery sequence position	Pseudoknot structure				Significance (A) and stability (B) scores of the stem-loop	
		Folding region of stem-loop	Size of stem 2	Number of bases in loops		A	B
		S1	S2	L1	L2		
BLV	gag-pro, 1592	1600–1619	3	2	6	–4.0	–5.1
	pro-pol, 2127	2138–2206	3	2	8	–2.3	–2.6
	or		3	3	12		
	or		4	2	18		
HTLV-1	gag-pro, 1769	1776–1800	3	2	4	–9.0	–8.0
	or		3	1	21		
	pro-pol, 2295	2302–2344	8	8	20	–3.6	–3.2
HTLV-2	gag-pro, 1772	1780–1802	6	2	5	–6.1	–6.4
	pro-pol, 2278	2285–2327	8	6	22	–4.8	–4.5
MMTV	gag-pro, 3402	3400–3477	3	2	4	–6.4	–3.2
	or		3	5	4		
MPMV	pro-pol, 4211	4216–4296	3	2	6	–5.9	–3.5
	gag-pro, 2713	2714–2815	5	2	20	–4.1	–2.4
	or		3	4	15		
	pro-pol, 3621	3629–3712	3	2	17	–4.0	–3.7
SRV-1	or		3	3	32		
	or		3	3	36		
	or		3	4	41		
	gag-pro, 2328	2337–2355	6	1	12	–3.3	–7.6
HIV-1	pro-pol, 3237	3245–3328	7	2	37	–3.4	–4.0
	gag-pol, 1634	1643–1668	3	3	8	–5.7	–6.2
HIV-2	gag-pol, 1844	1845–1896	6	3	4	–3.4	–2.6
SIVmac	gag-pol, 1843	1845–1895	6	2	4	–5.0	–3.8
SIVagm	gag-pol, 1751	1761–1835	3	5	7	–3.3	–3.7
	or		3	5	12		
	or		3	5	21		
EIAV	gag-pol, 1727	1727–1772	4	2	31	–4.0	–4.2
VISNA	and		4	3	44		
	gag-pol, 1769	1769–1814	7	5	8	–2.2	–4.7
RSV	gag-pol, 2482	2484–2583	8	5	11	–5.5	–3.9
Gypsy	gag-pol, 2381	2389–2427	4	2	17	–2.8	–3.5
	or		3	2	32		
	or		4	2	4		
MIAP	gag-pol, 3081	3090–3153	7	4	5	–3.2	–2.7
17.6	gag-pol, 2390	2403–2416	3	1	30	–3.9	–5.8

<sup>a</sup> The RNA pseudoknot structure is composed of stem 1 (S1) and stem 2 (S2). Loop L1 crosses the deep groove and loop L2 the shallow groove in the pseudoknot. The significance and stability scores of the distinct stem-loop (S1) are listed columns A and B, respectively.

es. Among them, only one predicted stem-loop structure just downstream of the frameshift site in the retrotransposon 17.6 is a quite small hairpin (four base pairs [bp]). The highly stable and statistically more significant RNA secondary structures adjacent to the frameshift sites of HIV-1, HIV-2, and SIVmac are close and similar to those depicted by Varmus and coworkers [19, 40].

### *Putative RNA Pseudoknots Downstream of the Frameshift Sites of Retroviruses and Transposable Elements*

The possible RNA tertiary structure following the frameshift sites of RSV is displayed in Figure 2b. The tertiary interaction is quite stable in the predicted pseudoknot by a Watson-Crick base pairing (S2) of the sequence GCGCUACA in the loop of the distinct stem-loop structure (S1) with the complementary region UGUAGCGC downstream of the RNA secondary structure S1. In the S2, there are 8 bp. As the two helical segments stack, the sequences CUGUU (5 nt) in the loop and GAAGGUAACU (11 nt) in the single-stranded region adjoined to the S1 connect opposite strands by bridging in the deep groove (L1) or the shallow groove (L2), respectively. In this case, the size of the loop L1 is 5, and the size of L2 is 11 (see Table 1). In general, L1 crosses the major groove, and L2 crosses the minor groove. However, this may not necessarily be the case for all pseudoknot types. The predicted pseudoknot structural model of RSV is similar to the model computed by Brierley et al. [1].

The nine putative pseudoknots downstream of the “slippery” sequence in the gag-pol overlapping of HIV-1, HIV-2, SIVmac, SIVagm, EIAV, and VISNA retroviruses, as well as related retrotransposons gypsy, MIAP, and 17.6, are depicted in Figure 3. Similarly, Figures 4 and 5 delineate the 12 potential pseudoknots that immediately follow the frameshift sites of gag-pro and pro-pol of BLV, HTLV-1, HTLV-2, MMTV, MPMV, and SRV-1 retroviruses, respectively. Among them, the predicted RNA pseudoknots of HTLV-1 (pro-pol) and VISNA (gag-pol) presented here are like the computer-predicted foldings suggested by Brierley et al. [1].

All of these predicted pseudoknots consist of a distinct stem-loop structure of high thermodynamic stability and statistical significance and a tertiary interaction region of at least three Watson-Crick base pairs between the hairpin or interior

loops and a complementary single strand downstream of the distinct stem-loop structure. The structural features of these predicted pseudoknots show that at least 2 nt (except for SRV-1) or 3 nts cross the deep or shallow grooves in stacking two double helical stems, respectively (see Table 1). Particularly, the pseudoknots following the “slippery” sequences of pro-pol frameshifting of HTLV-1 and HTLV-2 are structurally very well conserved. They contain an identical 13 bp (S1) with a bulge of nucleotide C and eight compensatory base changes in the tertiary interaction (S2). In the eight compensatory base changes of the conserved tertiary interaction, five base transitions and three base transversions are observed. The occurrence of compensatory base changes throughout the homologous stem S2 of related sequences HTLV-1 and HTLV-2 is very remarkable and allows for the high conservation of the tertiary structures despite their sequence divergence. It seems likely, therefore, that the highly conserved RNA tertiary structure is an essential element of the frameshifting signal of HTLV-1 and HTLV-2. Inspection of the structures in Figures 3–5 indicates that not all stems in the pseudoknotted structures might necessarily be coaxially stacked.

### Discussion

Our results reveal that 21 RNA pseudoknots predicted from the primary sequence of the junction domains of gag-pol or gag-pro and pro-pol of retroviruses occur consistently just downstream from the putative retroviral frameshift sites. The stem-loop structures involved in the complex tertiary structures are both highly stable and significant relative to others that may occur within these junction domains. For these distinct stem-loops, the standardized stability and significance scores, which measure the thermodynamic stability and statistical relevance of these RNA secondary structures in the folding process, are mostly less than  $-3.1$ . We have previously reported that the distribution of the lowest free energies generated from the randomly shuffled sequences follows approximately a normal distribution [22]. Thus, the probability that the significance scores of the segment would be obtained by chance can be approximately determined using the table of the standardized normal distribution. The smaller the probability of occurrence of a particular significance score, the more significant the secondary structure of the real biological segment. Similarly,



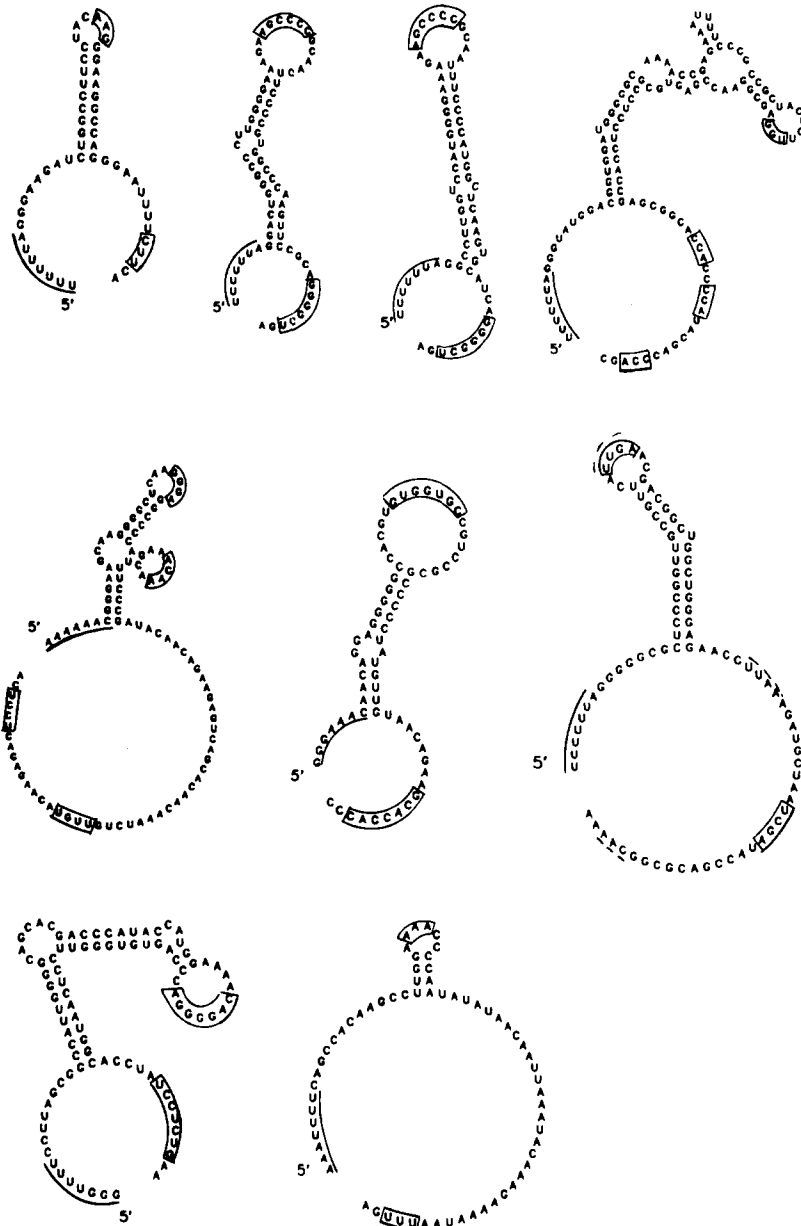


Figure 3. Potential RNA pseudoknot structures immediately 3' to the ribosomal (-1) frameshift sites in the gag-pol overlapping domains of nine retroviruses. First row: HIV-1 (HXB2), HIV-2 (ROD), SIVmac (K6W), and SIVagm (TYO-1). Second row: EIAV, VISNA, and gypsy. Third row: MIAP and 17.6. For the RNA pseudoknot of SIVmac, the first two base pairs in the stem-loop should be melted to form the quasi-continuous double-helical stem. The "slippery" sequence is underlined, and the bases that could participate in tertiary interactions (that is, stem 2 in Table 1) between the hairpin loop and downstream region are boxed. The bases involved in the alternative tertiary interactions are indicated with a dashed line.

lower stability scores correspond to a greater likelihood of a structure relative to alternate structures in the same sequence. Recently, we have also reported an empirical borderline [32] for assessing the stability and statistical significance of RNA secondary structures using a sample constructed from 20 randomly shuffled sequences having the same base composition as HIV-1 (BH10). The detailed statistical analysis of stability and significance scores from the sample shows that on the average, less than  $2.58 \pm 0.59$ , or  $0.35\% \pm 0.30\%$  of the observations have the stability scores less than  $-2.0$  or  $-3.0$ , and less than

$2.00 \pm 0.85$ , or  $0.14\% \pm 0.13\%$ , of the observations have significance scores less than  $-2.0$  or  $-3.0$ . The results presented in this study strengthen our previous finding [20] that stem-loop structures situated just 3' to the retroviral frameshift sites are both highly stable and statistically significant relative to others in the gag-pol or gag-pro and pro-pol junction domains (600 nt).

A pseudoknot is defined in this article as a base-pairing interaction of a loop with a free-base span (unpaired bases not within a loop) in an RNA structure. The pseudoknot structure may result in a quasi-continuous coaxial extension of stem re-

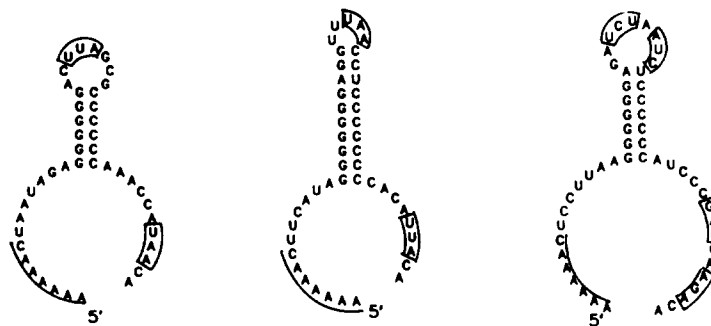
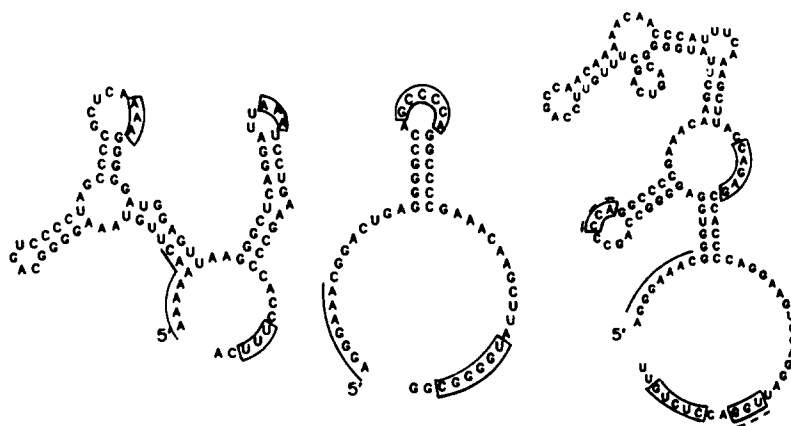


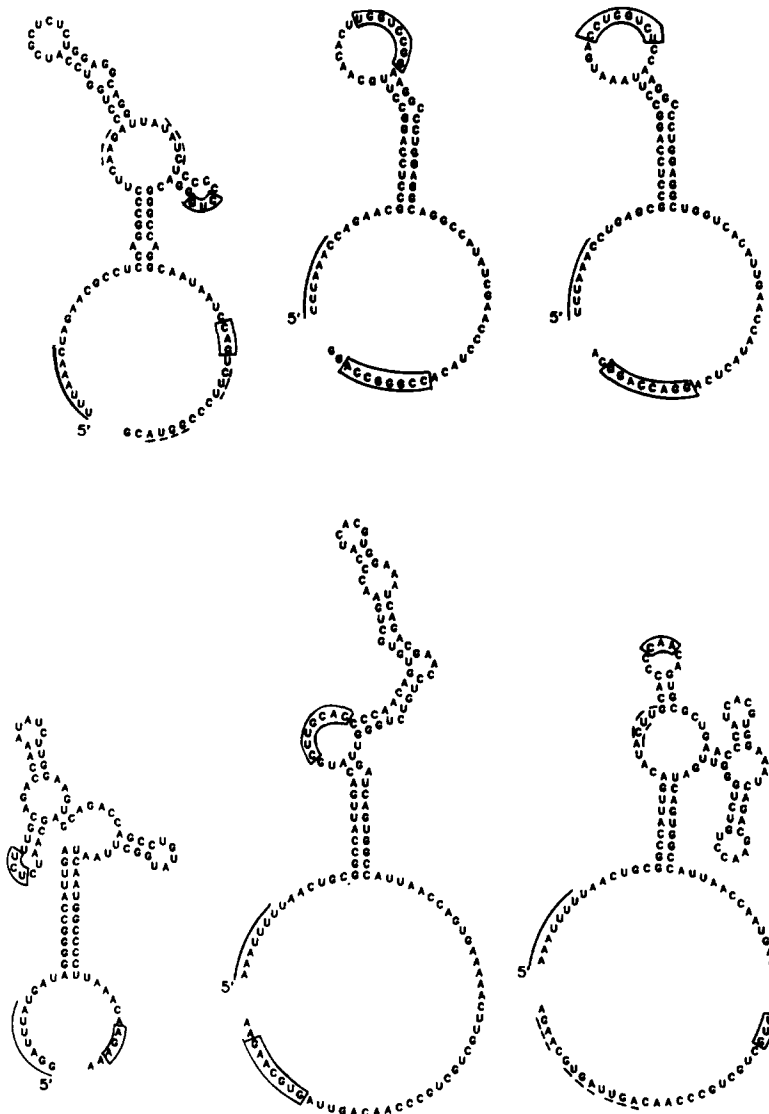
Figure 4. Potential RNA pseudoknot structures immediately 3' to the ribosomal (-1) frameshift sites in the gag-pro overlapping domains of six retroviruses. First row: BLV, HTLV-1, and HTLV-2. Second row: MMTV, SRV-1, and MPMV. For further details, see the caption to Figure 3.



gions by stacking two double helical segments (S1 and S2). Pleij et al. [12] have analyzed the constraints for the formation of the pseudoknot. The minimal length of S1 and S2 appears to be 3 bp, and the minimal sizes of the loops L1 and L2 that bridge the deep and shallow grooves appear to need 2 nt and 3 nt, respectively. Recently, they also proposed that the deep groove (L1) can be bridged over 6 bp by 1 nt only [14]. Among the predicted pseudoknots of these retroviruses, the RNA tertiary structure predicted in SIVmac seem likely to be doubtful according to the sterical constraints. In the predicted RNA tertiary structure adjacent to the "slippery" sequence of gag-pol frameshifting of SIVmac 4 nts cross the shallow groove of the double helix (18 bp) with a bulge of one base and an interior loop of two bases. The shortest distance that bridges a shallow or deep groove depends on the number of base pairs in the stems S1 and S2 of the pseudoknot; however, it is possible that other structural factors, the number and types of nucleotides in the bulge or interior loop, and their position in S1 are involved in the complex event. Recently, Lilley and coworkers [41] have found that bulges introduce pronounced

kinks into RNA helices. In the case of pseudoknots of SIVmac, the RNA tertiary structure seems sterically possible. One possible explanation is that the first 2 bp in S1 have to be melted to form the stacking of two stem segments. Thus, the stems S1 (16 bp) and S2 (6 bp) are connected by the loops L1 (2 nt) and L2 (7 nt) such that a right-handed, quasi-continuous double helix of 22 bp is formed. In this case, the distinct RNA stem-loop structure (S1) changes to be identical with the predicted structure model suggested by Madhani et al. [40].

It should be mentioned that Brierley et al. [1] have proposed six RNA pseudoknots of retroviruses RSV (gag-pol), BLV (pro-pol), HTLV-1 (pro-pol), MMTV (gag-pro and pro-pol), and VISNA (gag-pol). The putative RNA pseudoknots of retroviruses MMTV presented in the report are quite different from those structure models. We found that several alternative secondary structures can form in the downstream region of the frameshift site of the MMTV (gag-pro). Using the extensive Monte Carlo simulation, five distinct unusual folding regions adjacent to the frameshift site were detected. They are regions A (3409–



**Figure 5.** Potential RNA pseudoknot structures immediately 3' to the ribosomal (-1) frameshift sites in the pro-pol overlapping domains of six retroviruses. First row: BLV, HTLV-1, and HTLV-2. Second row: MMTV, SRV-1, and MPMV. For further details, see the caption to Figure 3.

3422), B (3409–3428), C (3425–3440), D (3452–3477), and E (3400–3477), where the sequence position of the last nucleotide C of the “slippery” sequence AAAAAAC is 3402. The significance and stability scores are, respectively, -4.46 and -3.18 for region A, -3.19 and -3.91 for region B, -3.80 and -3.81 for region C, -6.78 and -4.15 for region D, and -6.39 and -3.16 for region E. The RNA secondary structure predicted in region E (see Figure 5) includes the three hairpin loop structures predicted in regions A, C, and D. Obviously, the distinct stem-loop structure is appropriate for constructing the putative RNA pseudoknot of MMTV. The secondary structure entangled in the RNA pseudoknot suggested by Brierley et al. [1] is formed by folding region B. The hairpin loop structure is overlapped with the

hairpin loop structures (hairpins A and C) predicted in regions A and C. The sequences participating in potential tertiary base-pairing interactions in their model are involved in the two helical stems of the hairpins A and C. Thus, the alternate putative pseudoknot of MMTV is presented in the study.

The potential RNA pseudoknots were predicted by extensive searches of all possible interactions between the nucleotides of the loop in the distinct secondary structure detected above and the nucleotides downstream from the stem of the secondary structure. Recently, it has been verified that the RNA downstream of the (-1) frameshift site of IBV folds into a pseudoknot [1]. We found that the stem-loop structure involved in the RNA pseudoknot is highly stable and significant in the

junction region extending 300 nt upstream and downstream from the frameshift site of IBV. The distinct, unusual folding region can be detected by the extensive Monte Carlo simulation. The significance and stability scores are, respectively,  $-2.73$  and  $-4.73$ . The stem-loop structure was derived from the NEWFOLD program. The results of our search program applied to the data are depicted in Figure 6. The RNA pseudoknot of IBV proposed by Brierley et al. [1] can be predicted using the method presented in the report.

An interesting question raised by the study is why a retroviral frameshift site is closely followed by a pseudoknot structure consisting of coaxially stacked segments of a highly stable and statistically significant stem-loop structure, bridged by the connections of free nucleotides in the loop and downstream from the stem. Brierley et al. [1] have suspected that the tertiary RNA interactions may be a common feature of  $(-1)$  frameshift sites. In this study, we show that the occurrence of these complex tertiary structures is not random. Except for the stem-loop structure downstream of the BLV pro-pol junction region, the significance score or stability score of these distinct stem-loop structures by chance are all less than  $-3.1$ . These structures, being both highly stable and more significant within 300 nt upstream and downstream of the frameshift sites of these retroviruses, are apparently relevant to some biological function of these viruses. In the case of RSV, a mutational analysis [3] has demonstrated that a deletion just downstream from the stem-loop greatly inhibits retroviral frameshifting. If an additional 22 nt downstream were replaced, however, frameshifting was restored to approximately wild-type efficiency. This result is consistent with the unusual RNA pseudoknot detected just downstream from the frameshifting site of RSV (see Figure 2b). For the HTLV-1 and HTLV-2, the RNA pseudoknots adjacent to their "slippery" sequences of pro-pol frameshifting are structurally well conserved. The tertiary interactions in the conserved RNA pseudoknots involve a strict pattern of coordinate base changes, that is, a one-to-one correspondence between two positions where canonical pairing is maintained as a result of nucleotide substitutions. One possibility is that such extremely nonrandom RNA pseudoknot structures may act by stalling the translating ribosomes, where they promote the tRNA to slip back one nucleotide and pair with the codon in the  $-1$  frame [2]. Based on these results, the RNA

pseudoknot just downstream from the retroviral frameshifting site may be implicated in the  $(-1)$  ribosomal frameshifting event of retroviral translational systems. The possible control mechanism for frameshifting proposed by Jacks et al. [2, 3] and Brierley et al. [1] is strongly supported by this study.

However, a different mechanism of HIV-1 frameshifting have been proposed by Wilson et al. [42]. They showed that HIV frameshifting is mediated by a 26-nt sequence and is not dependent on the stem-loop structure following the frameshift site. Thus, the hypothesis that the signal of frameshifting in HIV is the UUUUUU was suggested [42]. HIV-1, HIV-2, and SIVmac are related lentiviruses. HIV-2 shares  $\sim 45\%$  nucleotide sequence similarity with HIV-1 and  $75\%$  sequence similarity with SIVmac [43]. The "slippery" sequence UUUUUUA at the frameshifting site is consistently conserved in these different primate immunodeficiency viral species and isolates. In this study, we show that the predicted RNA secondary structure entangled in the putative RNA pseudoknots are both highly stable and statistically significant relative to other possible structures in the gag-pol junction domains of HIV-1, HIV-2, SIVmac, and SIVagm. Using comparative sequence analysis, we observe that the secondary structural features occurring at and downstream of the gag-pol frameshift site are conserved in the 14 HIV-1 isolates and 7 HIV-2 and SIVmac isolates sequenced to date. Although the sequence shows an extreme variability in this region in different species and in independent viral isolates, the striking similarity in the complexity, overall appearance, and location of the stem-loop are consistently present in the 14 isolates of HIV-1 (data not shown) and 7 isolates of HIV-2 and SIVmac (Figure 7b).

Putative RNA pseudoknots in the 21 sequences are also analyzed by comparative sequence analysis. In seven RNA pseudoknots of HIV-2 and SIVmac retroviruses, the stable tertiary interaction of a Watson-Crick base pairs (S2) of the sequence AGCCCC in the hairpin loop with the complementary sequence GGGCU just downstream from the stem-loop (S1) is structurally very well conserved. In the 42 bp of the tertiary interactions of seven species, only one noncanonical base pair G-G is found in isolate FG of HIV-2. It seems likely that the structure conservation of the RNA pseudoknots is significant. Even though the G-G is considered a mismatch, the RNA tertiary structure of isolate FG of HIV-2 may still be formed by

## SEARCH RESULTS FOLLOW:

JOSHUA:&gt;bshapiro&gt;frame-shift&gt;pf7s.ZUK

- 13) C FREE BASE REGION  
 20) G INTERNAL LOOP-1  
 24) UGAGGCUCGG HAIRPIN LOOP-1  
 37) A INTERNAL LOOP-1  
 44) UUGCUAGUGG AUGUGAUCCU GAUGUUGUAA AGCGAGCCUU UGAUGUU FREE BASE REGION

NUMBER OF HELICES COMPUTED: 51

SORTED REGION LIST:

TOP 51 REGIONS

KEY SORTED ON: SECOND

(24 32 3 -3.8)	5' Half Region: UGA	3' Half Region: UCG
(27 45 2 -0.5)	5' Half Region: GG	3' Half Region: UU
(26 45 2 -0.5)	5' Half Region: AG	3' Half Region: UU
(31 46 3 -2.0)	5' Half Region: CGG	3' Half Region: UUG
(24 46 3 -1.0)	5' Half Region: UGA	3' Half Region: UUG
(27 48 3 -4.9)	5' Half Region: GGC	3' Half Region: GCU
(32 48 2 -1.5)	5' Half Region: GG	3' Half Region: CU
(26 48 2 -1.7)	5' Half Region: AG	3' Half Region: CU
(24 49 2 -0.7)	5' Half Region: UG	3' Half Region: UA
(28 51 3 -3.6)	5' Half Region: GCU	3' Half Region: AGU
(24 52 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(30 53 3 -3.4)	5' Half Region: UCG	3' Half Region: UGG
(29 53 2 -1.5)	5' Half Region: CU	3' Half Region: GG
(30 54 2 -2.3)	5' Half Region: UC	3' Half Region: GA
(31 56 2 -1.5)	5' Half Region: CG	3' Half Region: UG
(24 56 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(28 57 2 -1.9)	5' Half Region: GC	3' Half Region: GU
(24 58 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(30 59 3 -3.8)	5' Half Region: UCG	3' Half Region: UGA
(32 61 2 -1.3)	5' Half Region: GG	3' Half Region: UC
(27 61 2 -1.3)	5' Half Region: GG	3' Half Region: UC
(25 61 2 -2.3)	5' Half Region: GA	3' Half Region: UC
(32 62 2 -2.9)	5' Half Region: GG	3' Half Region: CC
(26 63 3 -4.6)	5' Half Region: AGG	3' Half Region: CCU
(27 63 2 -1.5)	5' Half Region: GG	3' Half Region: CU
(24 64 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(30 65 4 -5.3)	5' Half Region: UCGG	3' Half Region: CUGA
(31 67 2 -1.5)	5' Half Region: CG	3' Half Region: UG
(24 67 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(27 69 3 -2.4)	5' Half Region: GGC	3' Half Region: GUU
(26 69 2 -0.5)	5' Half Region: AG	3' Half Region: UU
(31 70 3 -2.0)	5' Half Region: CGG	3' Half Region: UUG
(24 70 3 -1.0)	5' Half Region: UGA	3' Half Region: UUG
(28 71 2 -1.9)	5' Half Region: GC	3' Half Region: GU
(24 72 2 -0.7)	5' Half Region: UG	3' Half Region: UA
(28 76 3 -5.1)	5' Half Region: GCU	3' Half Region: AGC
(24 77 2 -1.5)	5' Half Region: UG	3' Half Region: CG
(32 81 2 -2.9)	5' Half Region: GG	3' Half Region: CC
(32 82 2 -1.5)	5' Half Region: GG	3' Half Region: CU
* (25 83 8 -14.5)	5' Half Region: GAGGCUCG	3' Half Region: CGAGCCUU
(32 83 2 -0.5)	5' Half Region: GG	3' Half Region: UU
(26 84 3 -1.0)	5' Half Region: AGG	3' Half Region: UUU
(27 84 2 -0.5)	5' Half Region: GG	3' Half Region: UU
(24 85 5 -3.0)	5' Half Region: UGAGG	3' Half Region: CUUUG
(30 86 4 -4.3)	5' Half Region: UCGG	3' Half Region: UUGA
(31 88 2 -1.5)	5' Half Region: CG	3' Half Region: UG
(24 88 2 -0.5)	5' Half Region: UG	3' Half Region: UG
(27 90 3 -2.4)	5' Half Region: GGC	3' Half Region: GUU
(32 90 2 -0.5)	5' Half Region: GG	3' Half Region: UU
(26 90 2 -0.5)	5' Half Region: AG	3' Half Region: UU
(25 90 2 -0.5)	5' Half Region: GA	3' Half Region: UU

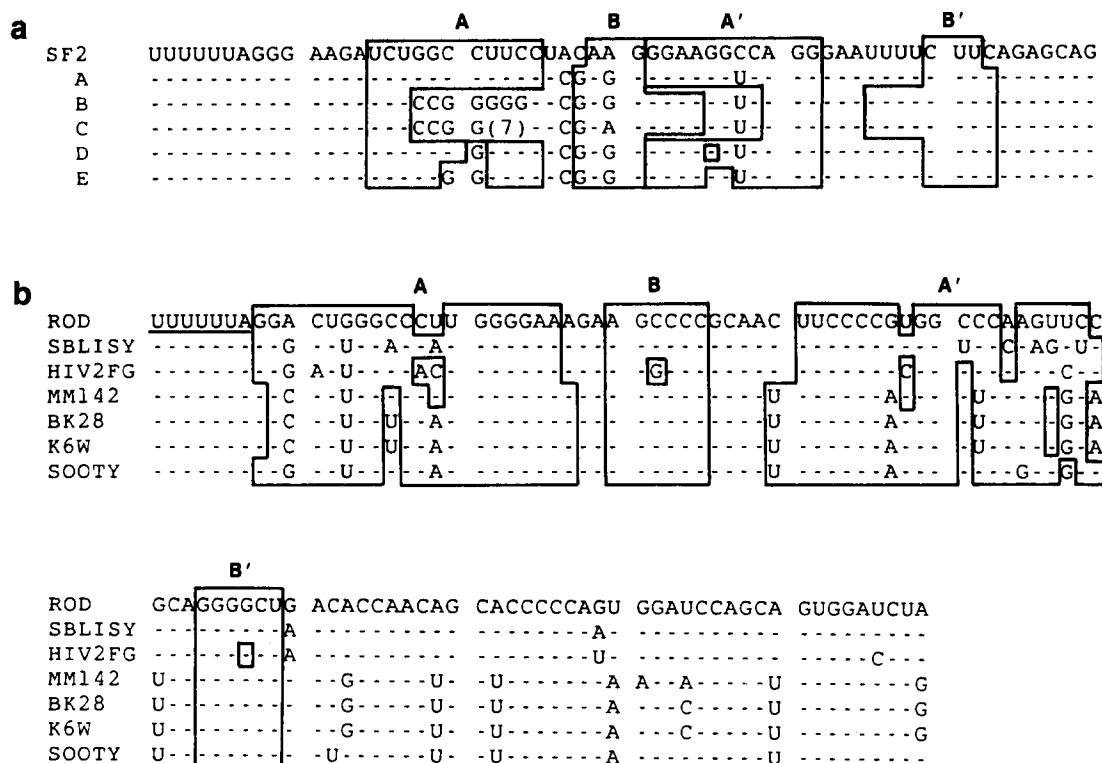


Figure 7. (a) Alignment of the sequences participating in putative tertiary interactions at the frameshift sites of the mutants used in the HIV-1 frameshifting experiment of Madhani et al. [40]. (b) Alignment of the sequences of distinct unusual RNA folding regions and additional 50 nucleotides in the downstream of three isolates of HIV-2 and four isolates of SIV. The "slippery" sequence is underlined, and stems 1 and 2 in the RNA pseudoknots are marked by boxes A and A', and B and B', respectively. Any bases within a box that are enclosed by another box fail to form an appropriate base pair.

at least three C-G base pairs in S2. Similarly, in the conserved RNA tertiary structure of 14 HIV-1 isolates, the sequence AAG in the hairpin loop forms normal Watson-Crick base pairing with a complementary sequence CUU 8 nt downstream

Figure 6. Output of a running search program applied to the IBV data presented in Brierley et al. [1]. The top part of the figure shows the individual loop or free-base segments and their 5' start position. The last sequence run represents that subsequence that is downstream from the predicted highly stable and significant structure. The bottom part of the output indicates the sorted quadruples with the half-region subsequences that form the potential pseudoknot stems. The annotation just above the list of quadruples indicates the total number of potential pseudoknot helices computed, the number that is listed, and the field (element of the quadruplet) on which the listing is sorted (in this case, the second field, the 3' position). Each quadruplet contains the 5' start position, 3' stop position, size, and energy of the potential knot regions. The potentially strongest tertiary interaction is located at the line marked by the asterisk (\*).

from the distinct stem-loop structure. However, one noncanonical base pair A-C is observed in isolates ELI and MN of HIV-1. Although a noncanonical base pair can also be considered a possible base pair in the RNA secondary or tertiary structure of tRNAs, ribosomal RNAs, and leader sequences of polioviruses [44-46], the tertiary interaction in these putative RNA pseudoknots of HIV-1 isolates is less stable than that predicted in HIV-2, SIV, and most other retroviruses. Moreover, the alternative tertiary interactions can also be found except for isolate HXB2, where the complementary sequence CUU is located 44 nt downstream from the distinct stem-loop structure. Thus the tertiary interactions in the putative RNA pseudoknot of HIV-1 are not absolutely conserved and less stringent. Nevertheless, our data demonstrate that the predicted RNA secondary structures adjacent to the frameshift sites of the 14 human immunodeficiency viruses are highly nonrandom and stable structures. Our results cannot explain why HIV frameshifting does not depend on stem-loop structures just downstream from the frameshift site.

Recently, Madhani et al. [40] have also observed that the frameshifting efficiency in the absence of the predicted RNA secondary structures at the frameshift site was not significantly different

from that determined for wild-type HIV-1. Interestingly, the RNA pseudoknots can be folded throughout these mutants, which are used to destabilize the RNA secondary structure downstream from the frameshift site of HIV-1 (isolate SF2) by Madhani et al. [40]. In the putative pseudoknot of the wild-type isolate SF2 of HIV-1, a quasi-continuous coaxial double helix of 14 bp is formed by stacking the stem S2 of 3 bp on S1 of 11 bp. The analysis of the potential RNA pseudoknots of these mutants shows that, even if the base mutations in the double-helical region S1 (see boxes A and A' in Figure 7a) destabilize the distinct, highly stable, and significant stem-loop structure (S1), the possible tertiary interaction between the bases in the hairpin loop and their complementary bases downstream from the stem-loop survive and are stabilized. For example, in the putative RNA pseudoknot of mutant B, S1 is destabilized by decreasing 8 bp, and the thermodynamic stability of tertiary interaction is strengthened by adding 5 bp to S2. Consequently, a quasi-continuous double helix of 11 bp is formed in the pseudoknot downstream of the frameshift site. Similarly, the quasi-continuous stems of 15, 11, 14, and 13 bp are also able to form in the pseudoknots of the mutants A, C, D, and E, respectively. Therefore, the thermodynamic stability of the complex tertiary structures can be partly compensated for by the strengthened tertiary interactions between the nucleotides in the hairpin loop and the free nucleotide span just downstream from the stem. If these pseudoknots indeed occur in these mutants, this may be a reasonable explanation for the experiments they did. Moreover, in the same report by Madhani et al. [40], the other experiments demonstrated that the feature of sequences following the frameshift site can have dramatic effects on the frameshifting of HIV-1.

In the absence of confirming experimental data, the prediction of RNA pseudoknot structure has some limitations. In this study, we focus on detecting thermodynamically highly stable and statistically significant RNA folding regions rather than specific predicted structures. Although the predicted pseudoknots must be verified, the identified distinct unusual folding regions just downstream of the frameshift site are apparently relevant to the ribosomal frameshift event of these viruses.

After this work was completed, we learned of an article by Ten Dam et al. [31] reporting results similar to ours. Both papers predict the potential

presence of pseudoknots downstream of the frameshift sites. This is gratifying, as substantially different methods were adopted. The method of Abrahams et al. [39] has the following features. It first enumerates a set of possible stems and then draws on these stems to construct a structure including stems that may form stable H-type pseudoknots. These pseudoknots consist of coaxial stacking of two stems with no intervening bases on one side [47]. This is similar to other kinetic secondary structure prediction programs [48, 49]. In contrast, our method is based on the evaluation of statistical significance, picking the most significant and stable structure. The validity of this approach has already been demonstrated in several experimental structures that have confirmed our predictions [20, 50, 51]. The fact that the two approaches yield similar results reinforces the theoretical predictions.

---

We thank Dr. M. Zuker for his interest in and support of our work.

---

## References

1. Brierley I, Digard, Inglis SC: *Cell* 57:537–547, 1989
2. Jacks T, Townsley K, Varmus HE, Majors J: *Proc Natl Acad Sci USA* 84:4298–4302, 1987
3. Jacks T, Madhani HD, Masiarz FR, Varmus HE: *Cell* 55:447–458, 1988
4. Rietveld K, van Poelgeest R, Pleij CWA, van Boom JH, Bosch L: *Nucleic Acids Res* 10:1929–1946, 1982
5. Rietveld K, Linschooten K, Pleij CWA, Bosch L: *EMBO J* 3:2613–2619, 1984
6. Joshi RL, Chapeville F, Haenni AL: *Nucleic Acids Res* 13:347–354, 1985
7. van Belkum A, Bingkum J, Rietveld K, Pleij CWA, Bosch L: *Biochemistry* 26:1144–1151, 1987
8. Dumas P, Moras D, Florentz C, Giege R, Verlaan P, van Belkum A, Pleij CWA: *J Biolmol Struct Dynam* 4:707–728, 1987
9. Clarke BE, Brown AL, Curry KM, Newton SE, Rowlands DJ, Carroll AR: *Nucleic Acids Res* 15:7067–7080, 1987
10. Goringe HU, Wagner R: *Nucleic Acids Res* 14:7473–7485, 1986
11. Deckman IC, Draper DE: *J Mol Biol* 196:323–332, 1987
12. Pleij CWA, Rietveld K, Bosch L: *Nucleic Acids Res* 13:1717–1731, 1985
13. Puglisi JD, Wyatt JR, Tinoco I Jr: *Nature* 331:283–286, 1988
14. Pleij CWA, Abrahams JP, van Belkum A, Rietveld K, Bosch L: In *Positive Strand RNA Viruses*. New York, Alan R Liss, 1987, pp 299–316
15. Tang CK, Draper DE: *Cell* 57:531–536, 1989

16. McPheeters DS, Stormo GD, Gold L: *J Mol Biol* 201:517-535, 1988
17. Moore R, Dixon M, Smith R, Peters G, Dickson C: *J Virol* 61:480-490, 1987
18. Rice NR, Stephens RM, Burny A, Gilden RV: *Virology* 142:357-377, 1985
19. Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE: *Nature* 331:280-283, 1988
20. Le S-Y, Chen J-H, Maizel JV Jr: *Nucleic Acids Res* 17:6143-6152, 1989
21. Le S-Y, Chen J-H, Currey KM, Maizel JV Jr: *Computer Applications in the Biosciences* 4:153-159, 1988
22. Le S-Y, Maizel JV: *Theor Biol* 138:495-510, 1989
23. Schwartz DE, Tizard R, Gilbert W: *Cell* 32:853-869, 1983
24. Myers G, Rabson AB, Josephs SF, Smith TF, Wong-Stall F: In *Human Retroviruses and AIDS*. Los Alamos, NM, 1989, pp
25. Sagata M, Yasunaga T, Tsuzuku-Kawamura J, Ohishi K, Ogawa Y, Ikiawa Y: *Proc Natl Acad Sci USA* 82:677-681, 1985
26. Sonigo P, Barker C, Hunter E, Wain-Hobson S: *Cell* 45:375-385, 1986
27. Power MD, Marx PA, Bryant ML, Gardner MD, Barr PJ, Luciw PA: *Science* 231:1567-1572, 1986
28. Saigo K, Kugimiya W, Matsuo Y, Inouye S, Yoshioka K, Yuki S: *Nature* 312:659-661, 1984
29. Marlor RL, Parkhurst SM, Corces VG: *Mol Cell Biol* 6:1129-1134, 1986
30. Meitz JA, Grossman Z, Lueders KK, Kuff EL: *J Virol* 61:3020-3029, 1987
31. Ten Dam EB, Pleij CWA, Bosch L: *Virus Gene* 2:121-136, 1990
32. Le S-Y, Chen J-H, Maizel JV Jr: In Sarma RH, Sarma MH (eds): *Structure and Methods, vol 1: Human Genome Initiative and DNA Recombination*. 1990, pp 127-136
33. Le S-Y, Malim MH, Cullen BR, Maizel JV Jr: *Nucleic Acids Res* 18:1613-1623, 1990
34. Chen J-H, Le S-Y, Shapiro B, Currey KM, Maizel JV Jr: *Computer Applications in the Biosciences* 6:7-18, 1990
35. Le S-Y, Chen JH, Chatterjee D, Maizel JV Jr: *Nucleic Acids Res* 17:3275-3288, 1989
36. Zuker M, Stiegler P: *Nucleic Acids Res* 9:133-148, 1981
37. Papanicolaou C, Gouy M, Ninio J: *Nucleic Acids Res* 12:31-44, 1984
38. Turner DH, Sugimoto N, Freier SM: *Ann Rev Biophys Chem* 17:167-192, 1988
39. Abrahams JP, Van Den Berg M, Van Batenburg E, Pleij CWA: *Nucleic Acids Res* 18:3035-3044, 1990
40. Madhani HD, Jacks T, Varmus HE: In Franza R, Cullen B, Wong-Staal F (eds): *The Control of HIV Gene Expression*. Cold Spring Harbor, NY, Cold Spring Harbor Laboratory, 1988, pp 119-125
41. Bhattacharyya A, Murchie AIH, Lilley DMJ: *Nature* 343:484-487, 1990
42. Wilson W, Braddock M, Adams SE, Rathjen PD, Kingsman SM, Kingsman AJ: *Cell* 55:1159-1169, 1988
43. Smith TF, Srinivasan A, Schochetman G, Marcus M, Myers G: *Nature* 333:573-575, 1988
44. Gutell RR, Weiser B, Woese CR, Noller HF: *Prog Nucleic Acid Res Mol Biol* 32:155-215, 1985
45. Douthwaite S, Powers T, Lee JY, Noller HF: *J Mol Biol* 209:655-665, 1989
46. Skinner M, Racaniello VR, Dunn G, Cooper J, Minor PD, Almond JW: *J Mol Biol* 207:379-392, 1989
47. Pleij CWA, Bosch L: In Dahlberg JE, Abelson JN (eds): *RNA Processing, Part A: General Methods, Methods Enzymol* 180:289-303, 1989
48. Dumas JP, Ninio J: *Nucleic Acids Res* 10:197-206, 1982
49. Martinez H: *Nucleic Acids Res* 12:323-334, 1984
50. Malim MH, Hauber J, Le S-Y, Maizel JV Jr, Cullen BR: *Nature* 338:254-257, 1989
51. Malim MH, Bohnlein S, Fenrick R, Le S-Y, Maizel JV Jr, Cullen BR: *Proc Natl Acad Sci USA* 86:8222-8226, 1989