VIRUS GENES 4:2, 121–136, 1990

RNA Pseudoknots: Translational Frameshifting and Readthrough on Viral RNAs

EDWIN B. TEN DAM, CORNELIUS W.A. PLEIJ, AND LEENDERT BOSCH
Department of Biochemistry, University of Leiden, Leiden, The Netherlands

Received November 17, 1989
Accepted January 29, 1990

Requests for reprints should be addressed to E.B. ten Dam, Department of Biochemistry, University of Leiden, Einsteinweg 5, P.O. Box 9502, 2300 CC Leiden, The Netherlands.

Key words: pseudoknots, ribosomal frameshifting, readthrough, RNA structure

Abstract

Ribosomal frameshifting on retroviral RNAs has been proposed to be mediated by slippage of two adjacent tRNAs into the −1 direction at a specific heptanucleotide sequence. Here we report a computer-aided analysis of the structure around the established or putative frameshift sites in a number of retroviral, coronaviral, toroviral, and luteoviral RNAs and two dsRNA yeast viruses. In almost all cases a stable hairpin was predicted four to nine nucleotides downstream of the shifty heptanucleotide. More than half of the resulting hairpin loops give rise to potential pseudoknotting with sequences downstream of this hairpin. Especially in the case of the shifty heptanucleotides U UUA AAC and G GGA AAC, stable downstream pseudoknots are present. Indications were also found for the presence of pseudoknots downstream of amber stop condons at readthrough sites in some retroviral RNAs.

Introduction

Translational frameshifting, although generally an abortive event during protein synthesis, is employed by various retroviruses to express the *pol* gene, encoding the reverse transcriptase and integrase. Frameshifting occurs at a defined site in the overlap region of the *gag* and *pol* genes and results in the synthesis of a *gag-pol* fusion protein (1,2). In some retroviral RNAs, e.g., mouse mammary tumor virus (MMTV) RNA (3,4), a double-frameshift event takes place, leading to the expression of a third reading frame encoding a protease. The ribosome is generally shifted into the −1 reading frame, but a shift into the +1 frame has been noted in one case for the related retroviral-like transposon Ty-1 (5,6). Studying the se-

quence requirements for ribosomal frameshifting during translation of Rous sarcoma virus (RSV) RNA, Jacks and coauthors found indications for a mechanism in which simultaneous slippage occurs of two adjacent ribosome-bound tRNAs by one nucleotide in the 5′ direction at the site of frameshifting. Comparison of the sequences at the known or suspected frameshift sites in reading frame overlaps of a number of retroviral RNAs revealed a consensus heptanucleotide, consisting of a run of three A, U, or G residues followed by the tetranucleotide UUUA, UUUU, or AAAC (7).

Interestingly, in the case of RSV RNA, the presence of such a shifty heptanucleotide appeared to be insufficient; an additional 147 nucleotides downstream of the frameshift site were also necessary for efficient frameshifting. Evidence was obtained that the additional 147 nucleotides in RSV RNA harbor a stable stem-loop structure. Furthermore, deletion analysis revealed that, beside this stem-loop structure, a downstream stretch of 20 nucleotides is essential (7). The authors suggested that these nucleotides may be involved in the formation of a pseudoknot.

However, for the human immunodeficiency virus (HIV-1), the stem-loop structure downstream of the frameshift site is dispensable, and efficient frameshifting can be mediated by a short sequence of 16 nucleotides around the frameshift site only. It was suggested that "retroviruses may divide in two broad classes, one using linear 'shifty' sequences (e.g., HIV) and the other using more elaborate mechanisms based on RNA secondary structure (e.g., RSV)" (8).

In this context it seemed of interest to examine the nucleotide sequences harboring frameshift sites in various overlap regions in more detail and to search for possibly tertiary interactions. Here we report the results of such a search performed with the computer. Beside the stable stem regions just downstream of the suspected frameshift sites already proposed (4,9,10), we also find strong indications for pseudoknotted structures downstream of the shifty heptanucleotide in more than half of the overlap regions examined, including those present in coronaviral, some plant viral RNAs, and a yeast dsRNA virus. During the course of this work, strong experimental evidence was reported for the presence of a pseudoknotted structure downstream of the frameshift site in infectious avian bronchitis virus (IBV) RNA. This pseudoknot was shown to be essential for efficient frameshifting of the ribosome in the orf1a/orf1b overlap region (11). These authors also proposed similar pseudoknots in a number of retroviral RNAs, some of which are identical to the ones resulting from our analysis.

An analysis of the region downstream of the site, where in some retroviral RNAs efficient readthrough of an amber stop codon occurs, also revealed potential pseudoknotted structures.

Methods

The overlap regions containing established or putative frameshift sites as tabulated (7) and 17 that have not been discussed before were folded in secondary

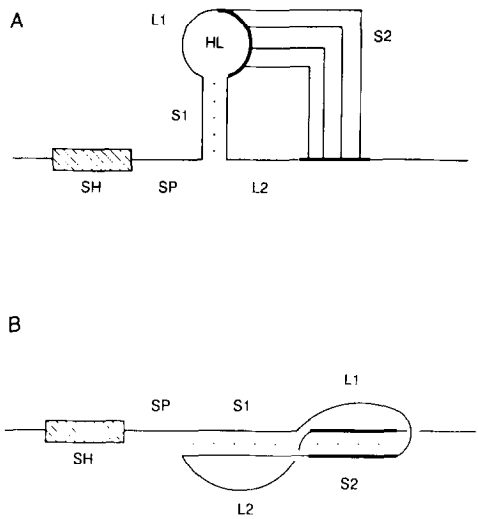

Fig. 1. Schematic representation of the structural elements of the frameshift region in viral RNAs. *A:* SH = shifty heptanucleotide; SP = spacer; S1 = stem 1; L1 = connecting loop 1; HL = hairpin loop; S2 = stem 2; L2 = connecting loop 2; see terminology section. *B:* Diagram illustrating the coaxial stacking of double helical segments S1 and S2.

structures using a program developed in our laboratory by Abrahams et al. (manuscript submitted for publication). This program is able to predict pseudoknotted structures involving hairpin loops, also coined *H-type pseudoknots* (12), and has been successfully applied for the prediction of a number of consecutive pseudoknots in the 5′ noncoding region of foot-and-mouth disease virus (FMDV) RNA (13) and of pseudoknots in various other viral RNAs (Pleij, unpublished observations). Stretches of about 250 nucleotides surrounding the shifty heptanucleotide were analyzed. Because only stem-loop structures downstream of the shifty sequence appear to be important (7,11), we have focused mainly on the sequence at the 3′ side of the frameshift site. Some pseudoknots involving bulge loops or multibranched loops are not predicted by this program. We, therefore, searched for these structural elements by visual inspection of the sequences if the proper stem-loop structures were found, taking into account the rules that are imposed by the geometry of the RNA-A double helix. The characteristics of RNA pseudoknots and their prediction and detection have been reviewed (12,14).

Results

Terminology

To simplify the description of the various structural elements around the frameshift region we introduce the terminology given in Fig. 1A. Essential fea-

tures are the heptanucleotide sequence SH, where the frameshift takes place, and the stem region S1 separated from SH by the spacer SP. If pseudoknotting involves a simple hairpin loop, as depicted in Fig. 1A, it is fully defined by the connecting loops L1 and L2 and the other stem region or "tertiary interaction" S2. Except for SH and SP, the symbols are derived from the nomenclature previously used to describe pseudoknots (12,14). Figure 1B shows a schematic presentation of the structure obtained after coaxial stacking of stem segments S1 and S2. A characteristic feature of the relatively simple pseudoknot illustrated in Fig. 1A is that the hairpin loop sequence participating in the tertiary interaction borders directly on the stem region of the hairpin. We call this type of pseudoknot *H*. When examining the retroviral structures we have not restricted ourself to a search for this particular type only, but have included pseudoknots that meet the more general definition: a structural RNA element formed upon basepairing of nucleotides within a loop with nucleotides outside that loop (12). An example of a more complicated pseudoknot is the one proposed for RSV RNA (see below).

Potential pseudoknots downstream of frameshift sites in retroviral RNAs

Table 1 presents the results of a computer-aided examination of 38 overlap regions harboring established or putative frameshift sites (7). We have included a number of sites from the retro-, luteo-, corona-, and toroviral groups and two yeast viruses not discussed before. For nearly all sequences tested, the computer program predicted a very stable hairpin, starting four to nine nucleotides downstream of SH, in agreement with observations by others for RSV, MMTV, HIV-1, HIV-2, and simian immunodeficiency virus (SIV_{MAC}) (4,9). Similar results were reported using a different RNA secondary structure-predicting program (11). This particular stem S1 was the most stable stem present in the 250 nucleotides surrounding SH, except for the luteovirus barley yellow dwarf virus (BYDV) (15) and the transposable element gypsy (16). In the latter two cases, it was the second best. In mouse intracisternal A particle (mouse IAP), such a hairpin is found if a G-A mismatch in S1 is allowed (17). S1 was neither predicted by the program nor found by eye for the transposable element 17.6 (11,18) and for the retrovirus SIV_{AGM} (10). The latter has U UUU UUA as the shifty sequence, and the absence of this hairpin is not surprising in view of the experimental results obtained for HIV-1 (8) (see Discussion). For HIV-2 and SIV_{MAC} we have included for S1 the stems as proposed earlier (9).

Interestingly, pseudoknotted structures were predicted directly by the program. A typical result obtained for the *gag-pro* overlap of SAIDS retrovirus-serotype 1 (SRV-1) RNA (20) is shown in Fig. 2A. The pseudoknot predicted here is of the H type (see Terminology), which is frequently observed in the noncoding regions of a number of other viral RNAs (21,22). The size of the connecting loops L1 and L2 meets the steric demands that result from stacking the two consecutive double-helical segments and forming a quasi-continuous helix. Accordingly, the single A

Table 1. Structural features downstream of established and possible frameshift sites in viral RNAs

RNA[a]	Overlap	SH[b]	SP[b]	S1[b]	HL[b]	S2[b]	L1[b]	L2[b]
BLV (31,32)	*gag-pro*	AAAAAAC	7	6	8	3	5	4
EIAV (27)	*gag-pol*	AAAAAAC	9	6	7	4	3	12
HTLV-I (28)	*gag-pro*	AAAAAAC	6	10	5	—	—	—
HTLV-II (30)	*gag-pro*	AAAAAAC	7	9	5	—	—	—
MMTV (3,4)	*gag-pro*	AAAAAAC	7	5	9	7	1	8
STLV-I (29)	*gag-pro*	AAAAAAC	6	10	5	—	—	—
RSV (36)	*gag-pol*	AAAUUUA	6	9	71[d]	8	8	17
17.6 (18)	*gag-pol*	AAAUUUU	—	—	—	—	—	—
MPMV (24)	*pro-pol*	AAAUUUU	7	9	66[d]	—	—	—
SRV-1 (20)	*pro-pol*	AAAUUUU	7	9	66[d]	—	—	—
SRV-2 (23)	*pro-pol*	AAAUUUU	7	9	66[d]	—	—	—
MMTV	*pro-pol*	GGAUUUA	4	12	57[d]	8	40[d]	14
BWYV (42)	orf2-orf3	GGGAAAC	6	5	6	4	2	6
MPMV	*gag-pro*	GGGAAAC	7	6	7	6	1	12
SRV-1	*gag-pro*	GGGAAAC	7	6	7	6	1	12
SRV-2	*gag-pro*	GGGAAAC	7	6	7	6	1	12
Visna (33)	*gag-pol*	GGGAAAC	6	7	19	7	5	14
SMRV-H[I] (26)	*gag-pro*	GGGAAAC	7	6	8	—	—	—
FIV (25)	*gag-pol*	GGGAAAC	8	5	8	6	2	11
SMRV-H[II]	*gag-pro*	GGGCCCC	4	5	6	5	1	9
Yeast A-L (45)	orf1-orf2	GGGUUUA	4	14	8	7	1	11
Yeast L1 (46)	orf1-orf2	GGGUUUA	4	14	8	7	1	11
BYDV (15)	39K-60K	GGGUUUU	5	7	18[d]	7	7	18
mouse IAP (17)	*gag-pol*	GGGUUUU	5	6	5	3	2	5
BLV	*pro-pol*	UUUAAAC	6	9	25	8	10	22[d]
HTLV-I	*pro-pol*	UUUAAAC	6	13	16	10	6	20
HTLV-II	*pro-pol*	UUUAAAC	6	13	16	8	6	22[d]
IBV (10)	orf1a-orf1b	UUUAAAC	6	11	8	7	1	32[d]
MHV A-59 (40)	orf1a-orf1b	UUUAAAC	5	11	13	11	2	30[d]
STLV-I	*pro-pol*	UUUAAAC	6	13	16	10	6	20
$PLRV_{WAG}$ (43)	orf2-orf3	UUUAAAC	6	4	6	4	2	9
BEV (40)	orf1a-orf1b	UUUAAAC	5	11	9	5	4	69[d]
$PLRV_{SCOT}$ (44)	orf2a-orf2b	UUUAAAU/C	6	4	6	4	2	8
gypsy (16)	*gag-pol*	UUUUUUA	7	15	8	—	—	—
HIV-1 (34)	*gag-pol*	UUUUUUA	7	11	5	—	—	—
HIV-2 (37)[c]	*gag-pol*	UUUUUUA	4	13	14	—	—	—
SIV_{MAC} (38)[c]	*gag-pol*	UUUUUUA	3	16	12	6	2	7
SIV_{AGM} (19)	*gag-pol*	UUUUUUA	—	—	—	—	—	—

[a] References for nucleotide sequences are indicated in parenthesis; for SMRV-H, I and II indicate the first and second hairpin downstream of SH, respectively; see also Fig. 2B.

[b] For definition see terminology section and Fig. 1. SP, HL, L1, and L2 are given in number of nucleotides; S1 and S2 in number of base pairs.

[c] Secondary structure as proposed (9).

[d] Presence of substructure in HL, L1, or L2, respectively.

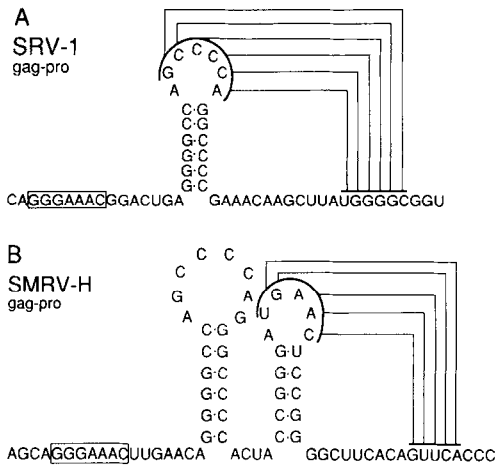

Fig. 2. Proposed structure around the frameshift site in the *gag-pro* overlap region of two type-D retroviral RNAs. A. SRV-1. The box indicates SH. B. SMRV-H. The box indicates SH. The stem of the 5′ proximal hairpin harbors the sequence G GGC CCC, which forms a potential second frameshift site.

residue is sufficient for crossing the deep groove of this helix over 6 bp, comparable to the pseudoknotting in the leader of the gene 32 mRNA of bacteriophage T4 (22).

Folding of the corresponding sequence in the *gag-pro* overlap of the closely related SRV-2 and Mason-Pfizer monkey virus (MPMV) RNA (23,24) yields a fully identical pseudoknot (not shown), which may reflect its functional importance. The sequence conservation in all three viral RNAs is absolute, however, which means that covariations in the stem regions, which generally provide support for the proposed structures, are lacking here. The *gag-pol* overlap region of feline immunodeficiency virus (FIV) contains a similar structure (25).

Inspection of the sequence in the corresponding overlap region in another type D retrovirus, SMRV-H (26), suggests a SH, as indicated in Fig. 2B. The SH is located at the same position as can be concluded unambiguously from a sequence alignment. Its sequence is identical to that of the three other type D retroviral *gag-pro* SHs mentioned above. The length of its SP is seven nucleotides, and its hairpin shows a strong resemblance to that of the other three related retroviruses (e.g., SRV-1 Fig. 2A). However, the formation of a H-type pseudoknot is not possible anymore due to the G insertion in HL and a U-to-C substitution in the complementary sequence downstream of the hairpin. Surprisingly, S1 now harbors a potential second SH, G GGC CCC, which in turn is separated by a SP of four nucleotides from the ideal H-type pseudoknot predicted by the program (Fig. 2B). This possible second frameshift site might compensate for the loss of the pseudoknot after the first site. Whether G GGC CCC can function as a SH sequence obviously remains to be seen, let alone that a second frameshift site indeed

is active in SMRV-H RNA. A similar situation was found in both SIV RNAs (see below).

The formation of a H-type pseudoknot is also possible in equine infectious anemia virus [EIAV (27), not shown], and another example is provided by human T-cell leukemia virus (HTLV-I) RNA in the *pro-pol* overlap (28). In this case, both S1 and S2 are exceptionally long. A stretch of ten nucleotides from the 16-membered loop is complementary to a sequence 20 nucleotides downstream of S1. The related simian T-cell leukemia virus (STLV-I) has a G-U pair in S2 substituted for an A-U pair (29). Much more substitutions are present in HTLV-II (30). S1 appears to be completely conserved, but various substitutions are found in L1 and L2. The three substitutions at the 3′ side of HL give rise to two mismatches, thereby shortening S2 and possibly interrupting the coaxial stacking on S1.

Similar deviations from what might be called the ideal H-type pseudoknot are encountered with a number of other retroviral RNAs, such as MMTV (*gag-pro*), bovine leukemia virus (BLV) (*pro-pol*) (31,32), and Visna virus (Visna) (33) RNA (not illustrated; see Table 1 and ref. 11).

A number of hairpins downstream of SH have a HL consisting of less than six nucleotides and in fact are not suitable for pseudoknotting [see HIV-I RNA [8,34,35], HTLV-I (*gag-pro*) and HTLV-II (*gag-pro*)]. The hairpin in the transposable element gypsy contains an eight-membered HL, but no possible pseudoknotting could be detected.

Non-H type pseudoknots can be more difficult to identify. An example is the one proposed for RSV RNA (36). The program predicted the same secondary structure downstream of the frameshift site as proposed (7). We assume, however, that the bottom part of S1 does not play a role in the frameshifting event for reasons outlined in the discussion. Visual inspection of the resulting hairpin revealed the potential pseudoknotting, as already described (11). In our view, this tertiary interaction can even be extended from 8 to 11 bp upon accepting the formation of a bulged U residue (not shown).

Retroviral RNAs lacking a pseudoknot downstream of SH

The relative complexity of the structure in RSV RNA is apparently not restricted to this RNA, but is found in a number of other retroviral RNAs with a similar HL of 30–60 nucleotides, often involved in internal hairpin formation themselves. Since the computer program is unable to predict pseudoknots harboring such multibranched loops, hairpins have to be inspected by hand. In doing so, no potential pseudoknots could be detected in the case of MPMV (*pro-pol*) or the related SRV-1 (*pro-pol*). In the *pro-pol* overlap of MMTV, a stretch of eight nucleotides of the loop (AGCCUGUA) was found to be complementary to a region just downstream of the hairpin (UACAGGCU). The significance of this complementarity is doubtful, however, because part of the stretch in the loop is already involved in a small hairpin (results not shown).

The group of retroviral RNAs having an SH consisting of the heptanucleotide U UUU UUA (e.g., HIV-1) shows other complexities. Long and stable hairpins with a small HL were proposed for HIV-2 and SIV_{MAC} (9). We note here that in both cases the sequence AGCCCC, occurring in HL, is complementary to the sequence GGGGCU, seven and nine nucleotides downstream of the stem, respectively (37,38). The program, however, predicted alternative structures, probably due to the presence of a number of alternating G- and C-rich regions downstream of SH.

The results obtained with SIV_{MAC} and SIV_{AGM} are puzzling, since the structure prediction suggested in both viruses a potential second SH, located 32 nucleotides downstream of the U UUU UUA sequence. In SIV_{AGM} a second SH sequence, A AAU UUU, is present in the *gag* gene reading frame, while in SIV_{MAC} a potential SH, U UUC CCC, is found at exactly the same position. Both SH sequences are followed by a stem region after three nucleotides. The hairpin found downstream of U UUC CCC in SIV_{MAC} RNA is reminiscent of the one present in RSV RNA and in some other retroviral RNAs, and a potential pseudoknot interaction with the long, single-stranded region further downstream is possible (not illustrated). We note that the situation described here for both SIV viruses is analogous to the one described for SMRV-H RNA in Fig. 2B.

For a member of another retrovirus subfamily, human spumaretrovirus (HSRV), surprisingly, no SH could be found in the overlap region, although the arrangement of its *gag* and *pol* genes suggests a −1 frameshift (39). Furthermore, no stable hairpins, let alone pseudoknots like the ones found in the other overlap regions, were predicted.

Coronaviral and toroviral RNAs

Coronaviruses are plus-stranded RNA viruses having large single-stranded RNA genomes with replication strategies different from retroviruses. However, it was recently shown for the coronavirus IBV that the overlap of the two open reading frames (orf1a and orf1b) of the putative polymerase gene contains the shifty sequence U UUA AAC, which is followed downstream by a pseudoknotted structure (7,10,11). Site-directed mutagenesis clearly demonstrated that the pseudoknot is involved in the very efficient frameshifting (25–30%). The program predicts essentially the same hairpin S1 as proposed by Brierley and coauthors (11), but we propose a slightly different tertiary basepairing, which enables a better coaxial stacking of S1 and S2, and is typical for an H-type pseudoknot (Fig. 3). This proposal is further supported by a comparison with the possible pseudoknot in the corresponding region of the related coronavirus mouse hepatitis virus (MHV) strain A-59 (40). Covariations in both S1 and S2 already prove that the pseudoknot exists in both viral RNAs. This is especially clear for S2, where three of these covariations are found, including the G-A pair in MHV. Note that such a G-A pair also occurs in the otherwise perfect stem S1 in IBV. Moreover, the

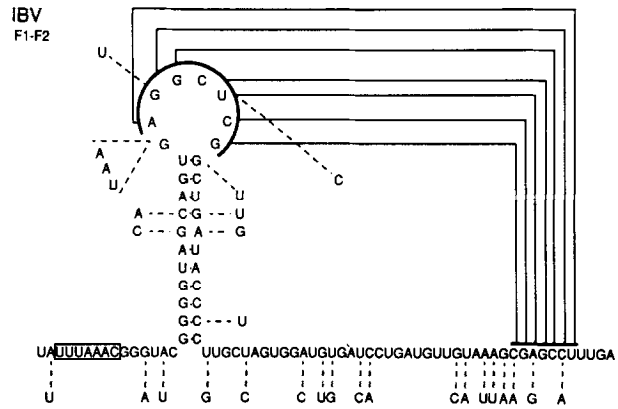

Fig. 3. Proposed structure around the frameshift site in the orf1a-orf1b overlap region of IBV. The box indicates SH. Dashed lines indicate base substitutions and the insertion of a UAA stop codon in MHV.

shortening of the MHV stem S1 at the top is compensated by the formation of an extra base pair (bp) in stem S2. It is further remarkable that the loop of the MHV hairpin shows an insertion of the stop codon triplet UAA, just in phase with the upstream orf1a coding region. This insertion extends stem S2 with another 2 or 3 bp (see Fig. 3). The single-stranded UG(U) stretch left may be just sufficient to cross the deep groove over 10–11 bp (14). We note that the 32-nucleotide-long connecting loop can be folded internally (not shown), but this does not interfere with the pseudoknotting itself. A similar structure is predicted for a member of the torovirus group, Berne virus (BEV) (40).

Luteoviral RNAs

Luteoviruses are plant viruses that have single-stranded plus-sense RNA genomes (41). Recently, the complete nucleotide sequences of three members of this group have been determined (42–44). The putative viral RNA-dependent RNA polymerase gene of BYDV is expressed by a −1 translational frameshift in the rather short overlap of 13 nucleotides. It was proposed that the UUUA just upstream of the stop codon signaled frameshifting, analogous to the phenomenon in some retroviruses and the coronavirus IBV (15). A possible stem-loop structure starting three nucleotides downstream from the UUUA sequence was also presented. We here propose that the SH is formed by the heptanucleotide G GGU UUU, followed after five nucleotides by this stem (15). Searching for possible pseudoknot formation by the 18-membered HL left open several possibilities for alternative, reasonable stable, secondary structures. A definitive proposal for the structure, therefore, cannot be offered.

More rewarding was the analysis of the sequence of beet western yellow virus

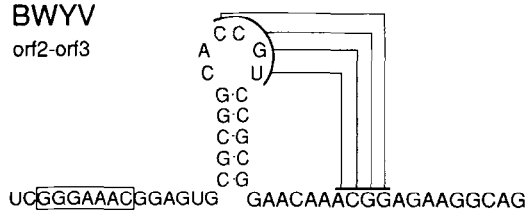

Fig. 4. Proposed structure around the frameshift site in the BWYV orf2-orf3 overlap region. The box indicates SH.

(BWYV) RNA (42). A potential candidate for a SH sequence in the right frame in the corresponding overlap region was found: G GGA AAC at position 1553 to 1559. It is followed after five nucleotides by a short but stable hairpin, which can form an H-type pseudoknot (Fig. 4). The nucleotide sequence of the closely related potato leaf roll virus (PLRV) RNA also has a potential SH. Its position [1662–1668 in $PLRV_{WAG}$ (43) or 1768–1774 in $PLRV_{SCOT}$ (44)] is identical to that of BWYV RNA, as can be concluded unambiguously from aligning both PLRV sequences with that of BWYV. Its composition, however, is rather different: U UUA AAU in $PLRV_{WAG}$ and U UUA AAU/C in $PLRV_{SCOT}$. Moreover, S1 in PLRV is shortened by 1 bp, but even more striking is the substitution of the U residue in HL of BWYV for a C in the Wageningen PLRV RNA sequence. This substitution weakens the pseudoknot structure, if existing at all. It is tempting to suggest that the pseudoknot requirement is relaxed because of the transition of SH from U UUA AAC to U UUA AAU (see Discussion).

Yeast viruses

L-A is a dsRNA virus of *Saccharomyces cerevisiae*. Its nucleotide sequence revealed two open reading frames, orf1 and orf2, overlapping by 130 bases. Orf2 is in the −1 reading frame with respect to orf1. A possible SH, G GGU UUA, is present in the overlap, followed after four nucleotides by a hairpin (45). Seven nucleotides of the eight-membered HL are complementary to a stretch of nucleotides that are 11 bases downstream of S1, thus again forming a potential pseudoknot. Another yeast dsRNA virus, L1, has an identical structure in the overlap region (46).

Readthrough of amber stop codons

Some retroviruses express their *pol* reading frame by suppressing an amber stop codon separating the *gag* and *pol* genes (47,48). A glutamine is inserted at this site, as shown in two cases (49,50). This very efficient suppression is caused by an

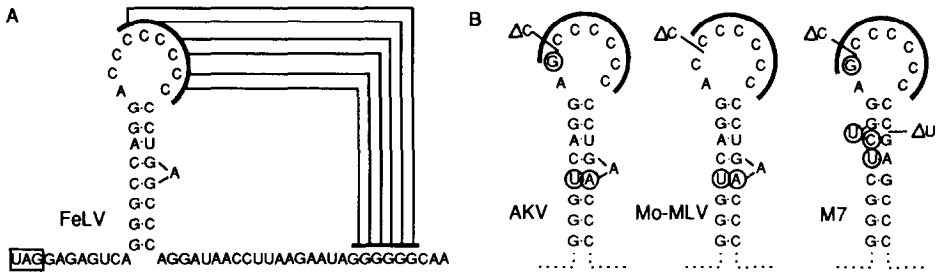

Fig. 5. Proposed structures downstream of the amber stop codon in retroviral RNAs using a suppression mechanism for the expression of the *pol* gene. A. FeLV. The box indicates the amber codon. B. Corresponding hairpins of AKV, Mo-MLV, and M7. Encircled residues indicate base changes with respect to FeLV. In these viruses HL is shortened by one C residue.

intrinsic *cis*-acting component of the viral RNA located within 300 nucleotides around the amber stop codon in AK virus (AKV) (51). A stable hairpin with the UAG codon in the loop was proposed as a secondary structure element that could play a role in the readthrough event (47,52). Recently, studies using site-directed mutagenesis around the *gag-pol* junction indicated that this stem-loop structure is important for virus activity (53). A similar hairpin is present in Moloney murine leukemia virus (Mo-MLV) (52). However, a role of this particular hairpin in the readthrough phenomenon is difficult to reconcile with the position of the amber codon in the loop region (see also 51). Moreover, the hairpin appears not to be conserved in M7 baboon endogenous virus (M7) and feline leukemia virus (FeLV, results not shown), nor in spleen necrosis virus (51). The computer program predicted other stable hairpins around the amber codon of FeLV and M7, which were not conserved in the other RNAs either. However, there is a structure motif that is conserved among all four viruses. We here note that one of the most stable hairpins possible in the entire FeLV genome occurs just downstream of the amber stop codon (54). This hairpin is capable of forming a pseudoknot. The loop contains a long stretch of only Cs at its 3′ side, which can form a very stable S2 with six G residues 18 nucleotides downstream of the hairpin (Fig. 5A). We emphasize the strong resemblance of this potential pseudoknot with some of those present in viral RNAs showing translational frameshifting (see above). Moreover, the distance from the UAG stop codon to S1 (eight nucleotides) reveals another striking resemblance (see also Discussion). Comparison of the FeLV sequence with those of three other retroviral RNAs, having established or putative suppressed amber codons (47,52,55), gives support to the proposed pseudoknot, though not in a decisive manner (Fig. 5B). Note again the single A residue in L1 of AKV and M7, and also of Mo-MLV, if C2255 indeed is a G residue, as was reported recently (53).

The nucleotide sequence of FeLV(54) in fact points to frameshifting as the mechanism of *pol* expression, because the *gag* and *pol* open reading frames are overlapping by five nucleotides, with *pol* in the +1 reading frame with respect to

gag. However, no signal for frameshifting can be found around the overlap region. No similarity between the 14-nucleotide sequence involved in +1 frameshifting in yeast Ty elements (5,6) is present upstream of the stop codon. Also no putative SH can be found. Removing 1 of the 10 consecutive C residues downstream of the amber codon enables a better alignment with the Mo-MLV sequence and would give this region of the FeLV genome an organization similar to that of the other three type-C retroviruses (49,50). This, and the presence of the pseudoknot in all four retroviruses discussed above, suggest that the FeLV expresses its *pol* gene by a readthrough mechanism.

Discussion

In this paper we have presented a computer-aided examination of the secondary structure and the potential pseudoknotting of the RNA region downstream of putative or established ribosomal frameshift sites of various viral RNAs. This search was inspired by the suggestion that a pseudoknot is involved in the frameshift event in the *gag-pol* overlap of RSV RNA (7). Our data, which include viral sequences not tabulated before, indicate that 26 of the 38 overlap regions studied here harbor potential pseudoknots. These pseudoknots are always found four to seven nucleotides downstream of a heptanucleotide sequence, where the translational frameshifting was demonstrated or supposed to take place. There are only three exceptions—EIAV, FIV, and SIV_{MAC}—where SP is nine, eight, and three nucleotides long, respectively. Some of the pseudoknotted structures found were of the same type as described previously for noncoding regions of viral RNAs, in which the stretch of nucleotides from HL basepairing with a complementary region outside this hairpin borders immediately on S1, enabling coaxial stacking of the two stem segments (pseudoknots of the H type, compare Fig. 1).

It is noteworthy that a stable hairpin downstream of the SH sequence was predicted for all overlap regions included in Table 1, except for the retrotransposon 17.6 and SIV_{AGM}. The failure to find a hairpin or pseudoknot in the overlap of the latter is consistent with the finding that the stable hairpin in HIV-1 RNA is dispensable for frameshifting (8). These authors proposed that two broad classes of retroviral RNAs exist, differing in their mechanism of frameshifting: one class using a short linear shifty sequence (like in HIV-1) and the other using RNA secondary structure for efficient frameshifting (e.g., RSV and IBV). In principle, the second class may be divided in two subclasses: one harboring a hairpin, the other a pseudoknot.

Our results suggest that a substantial, if not a major, part of the viral RNAs listed in Table 1 use a pseudoknotted structure for optimal shifting. A similar conclusion was recently presented by Brierley and coauthors (11), who reported that 14 out of 22 sequences examined appeared to contain the potential for pseudoknot formation. These authors also provided strong experimental evidence

that in IBV the pseudoknotted structure indeed is necessary for efficient frameshifting. It will be interesting to know if the same holds true for the majority of the viral RNAs analyzed here.

The question is, of course, what is it that makes a pseudoknot so suitable for inducing efficient frameshifting? We assume that it is not merely for formation of a structure more stable than a hairpin alone, because we were unable to find a correlation between the calculated stability of stem S1 (56) and the presence of a potential pseudoknot (results not shown). The number of base pairs in S1 was not found to be critical either. Two structural features distinguish an RNA pseudoknot from a classical RNA hairpin: the two connecting loops L1 and L2, of which the bases point into the deep groove and away from the shallow groove, respectively (57), and the quasi-continuity of the double helix. Which of these features induces the ribosome to shift into another reading frame remains to be established, however.

Another factor contributing to the extent of frameshifting could be the length of the spacer region, which varied between four and seven nucleotides. For IBV, changing SP from six to three or nine nucleotides, respectively, reduced or abolished frameshifting (11). Spacing between SH and the structure involved in frameshifting thus appears to be critical. In this respect it is striking that the distance between the amber stop codon and the pseudoknot at readthrough sites is almost equal to SP in frameshifting. The stem S1 as originally proposed for RSV RNA is in fact an exception, in that it starts immediately downstream of SH, forming a 14-bp stem, including a bulged C residue (7). We have chosen to disrupt 5 bp up to the bulged C residue, which leaves an SP of six nucleotides. The latter value is in the range of that of all the viral RNAs (see Table 1). Moreover, mutations in this six-nucleotide stretch did not alter the frameshift efficiency (7), which argues against the importance of the bottom part of the stem proposed.

No correlation between the SP size and the presence of a pseudoknot became apparent in the present comparison, however. A comparison of the sequences of the SH heptanucleotides is more suggestive (see Table 1). Two sequences stand out in overlaps harboring a pseudoknot: G GGA AAC and U UUA AAC. It is tempting to suggest, therefore, that the sequence AAAC, where the tRNA bound to the ribosomal A site is shifting, has to be followed by an elaborate RNA structure. The first three nucleotides of the SH heptanucleotide, however, play an additional role, as can be concluded from the group with the A AAA AAC sequence, which has some members for which potential pseudoknotting could not be established. An important factor to consider further may be the presence of C and G residues in the SH heptanucleotide, leading to more stable codon-anticodon interactions. In this case a longer stalling of the ribosome may be needed to increase the chance of the slippage event. Such a longer stalling may be achieved by an extra structural feature downstream of SH. In this respect, it is interesting to see that pseudoknotted structures may be involved as well in the efficient readthrough of an amber codon. It is conceivable that a common basis for both mechanisms is the need for a stalling of the translating ribosome, which is pro-

vided by pseudoknotted structures for reasons we do not yet know. If such a common basis is present, one can predict that a few changes in the nucleotide sequence around SH or the amber codon, respectively, could easily change a frameshifting viral RNA into one suppressing an amber codon and vice versa. However, first more information is needed about the actual requirement for amber stop-codon suppression of a downstream stem-loop structure or pseudoknot. The same holds true for a large number of viral RNAs having overlapping reading frames in which frameshifting occurs, despite the presently available data on RSV, HIV-1, and IBV.

Acknowledgments

We would like to thank Marianne Huisman, Mike Mayo, Wayne Gerlach, Peter Bredenbeek, and Willy Spaan for communicating data prior to publication; Jan Pieter Abrahams for making available the computer program; and Jan van Duin for stimulating discussions and reading the manuscript.

References

1. Jacks T. and Varmus H.E., Science *230*, 1237–1242, 1985.
2. Varmus H.E., Science *240*, 1427–1435, 1988.
3. Moore R., Dixon M., Smith R., Peters G. and Dickson C., J Virol *61*, 480–490, 1987.
4. Jacks T., Townsley K., Varmus H.E. and Majors J., Proc Natl Acad Sci USA *84*, 4298–4302, 1987.
5. Wilson W., Malim M.H., Mellor J., Kingsman A.J. and Kingsman S.M., Nucleic Acids Res *14*, 7001–7016, 1986.
6. Clare J.J., Belcourt M. and Farabaugh P.J., Proc Natl Acad Sci USA *85*, 6816–6820, 1988.
7. Jacks T., Madhani H.D., Masiarz F.R. and Varmus H.E., Cell *55*, 447–458, 1988.
8. Wilson W., Braddock M., Adams S.E., Rathjen P.D., Kingsman S.M. and Kingsman A.J., Cell *55*, 1159–1169, 1988.
9. Jacks T., Power M.D., Masiarz F.R., Luciw P.A., Barr P.J. and Varmus H.E., Nature *331*, 280–283, 1988.
10. Brierley I., Boursnell M.E.G., Binns M.M., Bilimoria B., Blok V.C., Brown T.D.K. and Inglis S.C., EMBO J *6*, 3779–3785, 1987.
11. Brierley I., Digard P. and Inglis S.C., Cell *57*, 537–547, 1989.
12. Pleij C.W.A. and Bosch L. in Dahlberg J.E. and Abelson J.N. (eds.) *Methods in Enzymology. RNA Processing. Part A. General Methods*. Academic Press, San Diego, 1989, Vol. 180, pp 289–303.
13. Clarke B.E., Brown A.L., Newton S.E., Rowlands D.J., Carrol A.R. and Currey K.M., Nucleic Acids Res *15*, 7067–7080, 1987.
14. Pleij C.W.A., Rietveld K. and Bosch L., Nucleic Acids Res *13*, 1717–1731, 1985.
15. Miller W.A., Waterhouse P.M. and Gerlach W.L., Nucleic Acids Res *16*, 6097–6111, 1988.
16. Marlor R.L., Parkhurst S.M. and Corces V.G., Mol Cell Biol *6*, 1129–1134, 1986.
17. Mietz J.A., Grossman Z., Lueders K.K. and Kuff E.L., J. Virol *61*, 3020–3029, 1987.
18. Saigo K., Kugimiya W., Matsuo Y., Inouye S., Yoshioka K. and Yuki S., Nature *312*, 659–661, 1984.

19. Fukasawa M., Miura T., Kasegawa A., Morikawa S., Tsujimoto H., Miki K., Kitamura T. and Hayami M., Nature *333*, 457–461, 1988.
20. Power M.D., Marx P.A., Bryant M.L., Gardner M.D., Barr P.J. and Luciw P.A., Science *231*, 1567–1572, 1986.
21. Pleij C.W.A., Abrahams J.P., van Belkum A., Rietveld K. and Bosch L. in Brinton M.A. and Rueckert R.R. (eds.) *Positive Strand Viruses. UCLA Symposia on Molecular and Cellular Biology, New Series*. Alan R. Liss, New York, 1987, Vol. 54, pp. 299–316.
22. McPheeters D.S., Stormo G.D. and Gold L., J Mol Biol *201*, 517–535, 1988.
23. Thayer R.M., Power M.D., Bryant M.L., Gardner M.B., Barr P.J. and Luciw, P.A., Virology *157*, 317–329, 1987.
24. Sonigo P., Barker C., Hunter E. and Wain-Hobson S., Cell *45*, 375–385, 1986.
25. Talbot R.L., Sparger E.E., Lovelace K.M., Fitch W.M., Pedersen, N.C., Luciw P.A. and Elder J.H., Proc Natl Acad Sci USA *86*, 5743–5747, 1989.
26. Oda T., Ikeda S., Watanabe S., Hatsushika M., Akiyama K. and Mitsunobu, F., Virology *167*, 468–476, 1988.
27. Stephens, R.M., Casey J.W. and Rice N.R., Science *231*, 589–594, 1986.
28. Hiramatsu K., Nishida J., Naito A. and Yoshikura H., J Gen Virol *68*, 213–218, 1987.
29. Inoue J.I., Watanabe T., Sato M., Oda A., Toyoshima K., Yoshida M. and Seiki M., Virology *150*, 187–195, 1986.
30. Shimotohno K., Takahashi Y., Shimizu N., Gojobori T., Golde D.W., Chen I.S.Y., Miwa M. and Sugimura T., Proc Natl Acad Sci USA *82*, 3101–3105, 1985.
31. Sagata N., Yasunaga T., Tsuzuku-Kawamura J., Ohishi K., Ogawa Y. and Ikawa Y., Proc Natl Acad Sci USA *82*, 677–681, 1985.
32. Rice N.R., Stephens R.M., Burny A. and Gilden R.V., Virology *142*, 357–377, 1985.
33. Sonigo P., Alizon M., Staskus K., Klatzman D., Cole S., Danos O., Retzel E., Tiollais P., Haase A. and Wain-Hobson S., Cell *42*, 369–382, 1985.
34. Wain-Hobson S., Sonigo P., Danos O., Cole S. and Alizon M., Cell *40*, 9–17, 1985.
35. Madhani H.D., Jacks T. and Varmus H.E. in Franza R., Cullen B. and Wong-Staal F. (eds.) *The Control of HIV Gene Expression*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory, 1988, pp. 119–125.
36. Schwartz D.E., Tizard R. and Gilbert W., Cell *32*, 853–869, 1983.
37. Guyader M., Emerman M., Sonigo P., Claver F., Montagnier L. and Alizon M., Nature *326*, 662–669, 1987.
38. Chakrabarti L., Guyader M., Alizon M., Daniel M.D., Desrosiers R.C., Tiollais P. and Sonigo P., Nature *328*, 543–547, 1987.
39. Maurer B., Bannert H., Darai G. and Flügel R.M., J Virol *62*, 1590–1597, 1988.
40. Bredenbeek P.J., *Nucleic Acid Domains and Proteins Involved in the Replication of Coronaviruses*. PhD Thesis, Utrecht, 1990.
41. Casper R. in Koenig R. (ed.), *The Plant Viruses; Polyhedral Virions with Monopartite RNA Genomes*, Vol. 3, Plenum, New York, 1988, pp. 235–258.
42. Veidt I., Lot H., Leiser M., Scheidecker D., Guilley H., Richards K. and Jonard G., Nucleic Acids Res *16*, 9917–9932, 1988.
43. Van der Wilk F., Huisman M.J., Cornelissen B.J.C., Huttinga H. and Goldbach R., FEBS Lett *245*, 51–56, 1989.
44. Mayo M.A., Robinson D.J., Jolly C.A. and Hyman L., J Gen Virol *70*, 1037–1051, 1989.
45. Icho T. and Wickner R.B., J Biol Chem *264*, 6716–6723, 1989.
46. Diamond M.E., Dowhanick J.J., Nemerhoff M.E., Pietras D.F., Tu C, and Bruenn J.A., J Virol *63*, 3983–3990, 1989.
47. Shinnick T.M., Lerner R.A. and Sutcliffe J.G., Nature *293*, 543–548, 1981.
48. Valle R.P.C. and Morch M., FEBS Lett *235*, 1–15, 1988.
49. Yoshinaka Y., Katoh I., Copeland T.D. and Oroszlan S., Proc Natl Acad Sci USA *82*, 1618–1622, 1985.

50. Yoshinaka Y., Katoh I., Copeland T.D. and Oroszlan S., J. Virol *55*, 870–873, 1985.
51. Panganiban A.T., J Virol *62*, 3574–3580, 1988.
52. Herr W., J Virol *49*, 471–478, 1983.
53. Jones D.S., Nemoto F., Kuchino Y., Masuda M., Yoshikura H. and Nishimura S., Nucleic Acids Res *17*, 5933–5945, 1989.
54. Laprevotte I., Hampe A., Sherr C.J. and Galibert F., J. Virol *50*, 884–894, 1984.
55. Tamura T., J Virol *47*, 137–145, 1983.
56. Turner D.H., Sugimoto N. and Freier S.M., Ann Rev Biophys Biophys Chem *17*, 167–192, 1988.
57. Dumas P., Moras D., Florentz C., Giegé R., Verlaan P., van Belkum A. and Pleij C.W.A., J Biomol Struct Dyn *4*, 707–728, 1987.