The EMBO Journal vol.13 no.17 pp.4137–4144, 1994

Pseudoknot-dependent read-through of retroviral *gag* termination codons: importance of sequences in the spacer and loop 2

Norma M.Wills[1], Raymond F.Gesteland[1,2] and John F.Atkins[2]

[1]Howard Hughes Medical Institute and [2]Department of Human Genetics, 6160 Eccles Genetics Building, University of Utah, Salt Lake City, UT 84112, USA

Communicated by J.F.Atkins

Retroviruses whose *gag* and *pol* genes are in the same reading frame depend upon ~5% read-through of the *gag* UAG termination codon to make the gag–pol polyprotein. For murine leukemia virus, this read-through is dependent on a pseudoknot located eight nucleotides 3′ of the UAG. Other retroviruses whose *gag* and *pol* genes are in the same frame can potentially form similar pseudoknots 3′ of their UAG codons. Beyond the similar secondary structures, there is strong sequence conservation in the spacer region and in loop 2 of the pseudoknots. The detrimental effects of substitutions of several of these conserved spacer and loop 2 nucleotides in the murine leukemia virus sequence show their importance for the read-through process. The importance of specific nucleotides in loop 2 of the pseudoknot contrasts with the flexibility of sequence in loop 2 of the most intensively studied frameshift-promoting pseudoknot which occurs in infectious bronchitis virus. Two nucleotides in loop 2 of the murine leukemia virus pseudoknot, which were shown to be important by mutagenic analysis, display hypersensitivity to the single-strand specific nuclease, S1. They are likely to be particularly accessible or are in an unusually reactive conformation.

Key words: pseudoknot/read-through/recoding/retrovirus

Introduction

Stop codons in particular contexts in mRNAs have dual meanings: the majority of ribosomes terminate translation and the minority of ribosomes insert an amino acid and continue translation to yield a read-through polypeptide. For the UAG stop codon, the efficiency of read-through at these special sites is 1–5% which is ~1000-fold above background 'error' rates (Capone *et al.*, 1986). The amino acid inserted is glutamine whose normal codons are CAG and CAA (Yoshinaka *et al.*, 1985a,b). At other particular sites, UGA stop codons can be decoded as tryptophan. In a few mRNAs, the function of the UGA termination codon is different; it specifies selenocysteine (Böck *et al.*, 1991; Berry *et al.*, 1993). The sequence context of terminators dictates which have special meaning. In many cases, short adjacent sequences act as 'stimulators' to cause read-through (Skuzeski *et al.*, 1991; Tate and Brown, 1992; Li and Rice, 1993; for a review see Atkins *et al.*, 1990). However, complex RNA structures, i.e. pseudoknots, are known to be crucial for some recoding events involving frameshifting and read-through. Simple pseudoknots consist of two interdigitated stems which are coaxially stacked and joined by two loop regions which span the grooves of the helix (see Figure 1). For read-through of the UAG terminator of the *gag* gene of Moloney murine leukemia virus [MuLV (Philipson *et al.*, 1978; Yoshinaka *et al.*, 1985a)], a pseudoknot, located 3′ to the stop codon was predicted (ten Dam *et al.*, 1990) and shown to be required (Wills *et al.*,1991; Feng *et al.*, 1992). Read-through yields the gag–pol fusion polypeptide which is the only source of *pol* products, since there is no independent ribosome initiation at the start of the *pol* gene.

Evidence that a pseudoknot is the stimulatory element for MuLV read-through comes from *in vitro* translation experiments: mutations that disrupted base-pairing in either of the two stems of the pseudoknot greatly decreased or abolished read-through, whereas compensatory mutations that restored base-pairing led to recovery of read-through activity.

The ~60 nt 3′ of the UAG *gag* termination codon of several retroviruses whose *gag* and *pol* genes are in the same reading frame show considerable conservation. This group of viruses will be referred to as the read-through retroviruses. Included in this group are several murine retroviruses, Moloney, Friend, Abelson and AKV murine leukemia viruses, murine sarcoma virus [MSV (Reddy *et al.*, 1981; Van Beveren *et al.*, 1981)], and neurotropic murine retrovirus [CAS-BR-E (Perryman *et al.*, 1991)]. The *gag–pol* junctions of these murine virus are identical or differ by, at most, a few nucleotides. In addition, there are read-through retroviruses of feline, simian and avian origin: feline leukemia virus [FeLV (Yoshinaka *et al.*, 1985b)], baboon endogenous virus [BaEV (Kato *et al.*, 1987)], gibbon ape leukemia virus [GaLV (Delassus *et al.*, 1989)] and spleen necrosis virus [SNV (Weaver *et al.*, 1990)]. The regions of conserved sequence can be folded into strikingly similar pseudoknot structures (ten Dam

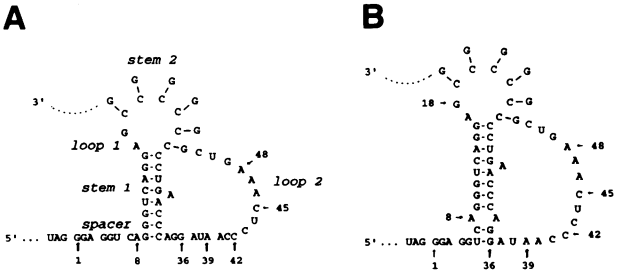

Fig. 1. RNA structures proposed to stimulate read-through at the *gag–pol* junction of MuLV. (**A**) RNA pseudoknot predicted by ten Dam *et al.* (1990). (**B**) Alternative RNA pseudoknot with extended base-pairing in stem 1 proposed by Felsenstein and Goff (1992).

et al., 1990; Hatfield *et al.*, 1992) (see Figure 2). The conservation of secondary structure in the region of the *gag* termination codon may be relevant to the read-through mechanism. Not only are the proposed stem regions very similar, but there are invariant nucleotides in the sequence between the UAG and the start of the pseudoknot and in loop 2. In the spacer region, U6 and C7 are found in all of the read-through retroviruses. G2 and G5 are highly conserved but some variations occur. Alterations in the spacer region can influence read-through in MuLV (Honigman *et al.*, 1991; Feng *et al.*, 1992) but have not been comprehensively analyzed. An extensive mutagenic analysis of the spacer nucleotides is presented here.

In the major single-stranded region of the pseudoknot, loop 2, several nucleotides are invariant among the read-through retroviruses, G35, G36, U38, A39, A40, U44 and U50, raising the possibility of their importance for the read-through process. This has been tested for MuLV and is reported here. The possible involvement of sequences in loop 2 of the MuLV pseudoknot contrasts sharply with what is known about loop 2 of the pseudoknot that stimulates frameshifting in infectious bronchitis virus (IBV) where drastic changes have no effect on frameshifting as long as a minimal size, eight nucleotides, is maintained to allow formation of the structure (Brierley *et al.*, 1991).

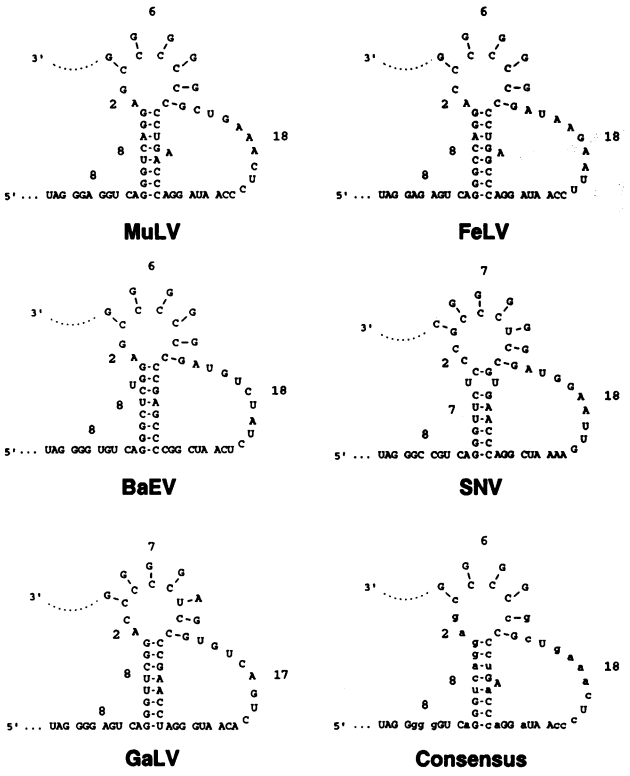

Fig. 2. Potential pseudoknot structures in the regions 3′ to UAG codons at the *gag–pol* junctions of several retroviruses. Shown are the sequences of Moloney murine leukemia virus, feline leukemia virus, baboon endogenous virus, spleen necrosis virus, gibbon ape leukemia virus and a consensus pseudoknot (including the sequences of other MuLVs, Friend, Abelson and AKV, murine sarcoma virus and neurotropic murine retrovirus, which are not shown). In the consensus pseudoknot, nucleotides that are conserved in all the viruses are shown in upper case, those that are found in the majority of the viruses are shown in lower case. (There are alternative structures possible for these sequences, for example, see Figure 1.)

Results

Potential pseudoknot sequences promote read-through

Complementary oligonucleotides were cloned in-frame into vector pRW201 between the coding regions of chloramphenicol acetyl transferase (CAT) and β-galactosidase (*lacZ*) as previously described (Wills *et al.*, 1991). Plasmid DNAs were transcribed with T7 RNA polymerase and the transcripts were translated in a reticulocyte lysate system. Read-through was measured by comparing the amounts of 73 kDa (read-through) and 42 kDa (termination) products. Read-through levels are reported as a relative percentage of the wild-type level (100%). The absolute level of wild-type read-through *in vitro* is 2–5% (Philipson *et al.*, 1978; Wills *et al.*, 1991; Feng *et al.*, 1992).

Oligonucleotides encoding the *gag–pol* junctions of FeLV, BaEV, GaLV and SNV were cloned and tested for read-through. These pseudoknot regions (see Figure 2) contained five codons 5′ and 19–20 codons 3′ of the UAG codon comparable with the wild-type MuLV construct previously tested (Wills *et al.*, 1991). All of the viral sequences tested promoted read-through (Figure 3). Experimental evidence supports the pseudoknot requirement for read-through of the UAG codon of MuLV *in vitro* (Wills *et al.*, 1991; Feng *et al.*, 1992). The effects of disruptive and restorative mutations in the proposed pseudoknots of FeLV, BaEV, GaLV and SNV have not been examined, therefore the requirement of the structure for read-through has not been established. However, the strong similarities between the sequences in this region of MuLV and the other viruses make it likely that all of the proposed pseudoknots are important for read-through. In support of this conclusion, read-through was increased in a SNV construct where the wild-type U-U mismatch

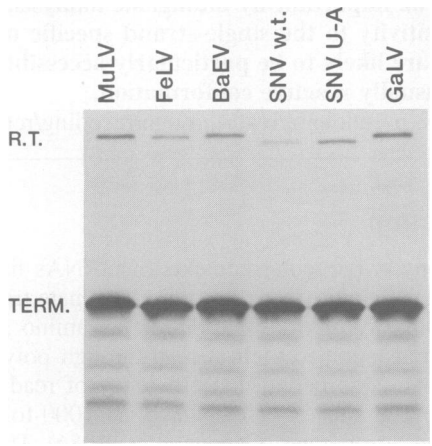

Fig. 3. The potential pseudoknot sequences of several retroviruses promote read-through. *In vitro* translations of transcripts containing the *gag–pol* junctions of the indicated viruses cloned between the open reading frames of CAT and β-galactosidase. The termination product (TERM.) is 42 kDa and the read-through product (R.T.) is 73 kDa. The lane labelled 'SNV U-A' contains a mutation in the SNV sequence, U27→A, resulting in the potential formation of an additional base pair in stem 1.

in stem 1 of the pseudoknot was replaced by A-U thereby increasing the stability of the stem (Figure 3). In another SNV construct, stem 2 was reduced by a single

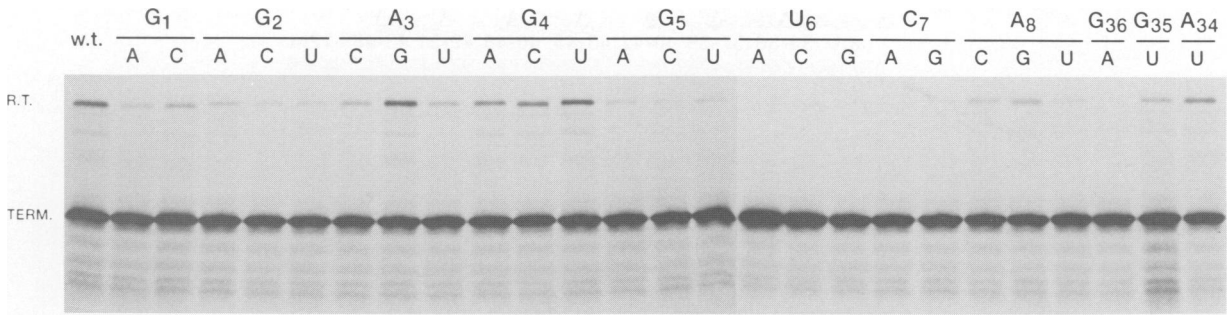

Fig. 4. Effects of mutations in the spacer region of MuLV. Nucleotides are numbered starting at the G residue following UAG. G1, G5, U6 and C7 are conserved among the read-through retroviruses. Termination and read-through products are labelled TERM. and R.T., respectively.

base pair and no read-through was detected (data not shown).

Mutations in the spacer

There is evidence that the conserved purine-rich sequence immediately 3′ to the UAG codon is involved in read-through. The effect of this sequence could be due to a local context effect (Honigman *et al.*, 1991; Feng *et al.*, 1992) and/or be exerted via secondary structure (Felsenstein and Goff, 1992). We undertook a more complete analysis of base substitutions at each position in this region, in the context of the wild-type MuLV pseudoknot, to address these questions. The results are shown in Figure 4.

Replacement of G_1, conserved among the read-through retroviruses, with C resulted in a reduction in read-through to ~50% of wild-type. However, replacement of G_1 by A decreased read-through to ~30%. These results contrast with those of Feng *et al.* (1992) who reported <5% read-through with either A or C substituted for G1. Because of the discrepancy between results with the G1→C constructs, another independent clone was isolated and it also gave 50% of the wild-type level, in agreement with our original result. (G1→U was not tested since it would create an in-frame UGA stop codon). Substitutions of G2, which is over-represented but not conserved in the read-through viruses, with A, C or U, gave read-through levels of 20–30% of wild-type. In contrast, substitution of A3 with G causes a 10% increase in read-through. However, replacement of A3 with a pyrimidine, C or U, lowers read-through to 40–50%. In all of the read-through retroviruses, a purine is found at the equivalent position of A3 in MuLV. G4 is not strongly conserved among the read-through retroviruses. As also shown in Figure 4, mutation of G4 to U had no effect on read-through, while substitution with A or C had marginal effects. G5, conserved but not invariant, when changed to A, C or U decreased read-through to 20–30%. U6 and C7 are invariant among the read-through retroviruses. Mutation of U6 to A, C or G, or of C7 to either A or G, drastically reduced read-through to <10% of wild-type. Mutation of A8, which is over-represented but not invariant, to C, G or U lowered read-through to 20–50%, with U having the greatest effect. These results show that the invariant residues in the spacer region have different effects on read-through. While U6 and C7 influence read-through dramatically, G1 and G5 have a lesser effect. Not surprisingly, there is greater tolerance for base substitutions of the less conserved residues, G2, A3, G4 and A8.

The requirement for specific nucleotides in the spacer region may be due to particular intermolecular interactions with protein or RNA, or, alternatively, to intramolecular interactions, i.e. base-pairing in a more complicated structure. The detrimental effects of all base substitutions for U6 or C7 may be due to disruption of base-pairing in the secondary structure proposed by Felsenstein and Goff (1992) and shown in Figure 1B where U6 pairs with G36 and C7 pairs with G35. To further investigate the possibilities of secondary structure involving these residues, mutations of A34, G35 and G36 were constructed. Mutation of A34→U, which would allow formation of an additional base pair in stem 1, had no significant effect on read-through (Figure 4). [In a construct of SNV, the potential of an additional base pair near the top of stem 1 causes a notable increase in read-through (Figure 3).] Mutation of G36→A would be expected to work as well as, or better than, wild-type if base-pairing is important. However, read-through was decreased to ~30% (Figure 4). Destabilization of stem 1 by mutations which disrupt base-pairing reduce read-through to <40% (Wills *et al.*, 1991). Changing G35 to U would be predicted to reduce read-through to a similar extent, but this construct retains ~60% of wild-type activity (Figure 4). These results taken together support the pseudoknot model in Figure 1A as opposed to the model in Figure 1B. Even though extended base-pairing in stem 1 is not supported by our data, there remains the possibility that nucleotides in the spacer region may interact with nucleotides elsewhere in the pseudoknot.

Mutations in loop 1

In the model for the MuLV pseudoknot, loop 1 consists of two nucleotides, A17 and G18. Loop 1 of a pseudoknot spans the major groove of the coaxially stacked stems (Pleij *et al.*, 1985). The identities of the two nucleotides in loop 1 of the pseudoknot that promotes frameshifting in IBV are not important (Brierley *et al.*, 1991). To find out whether this is the case for the MuLV pseudoknot, loop 1 residues were mutated and the effects on read-through were measured. A, C and U could substitute for G18 without effect (data not shown). U and C were acceptable replacements for A17, but substitution of G reduced read-through drastically (data not shown). The reduction in read-through may be attributable to competition for base-pairing with C24, proposed to base-pair with

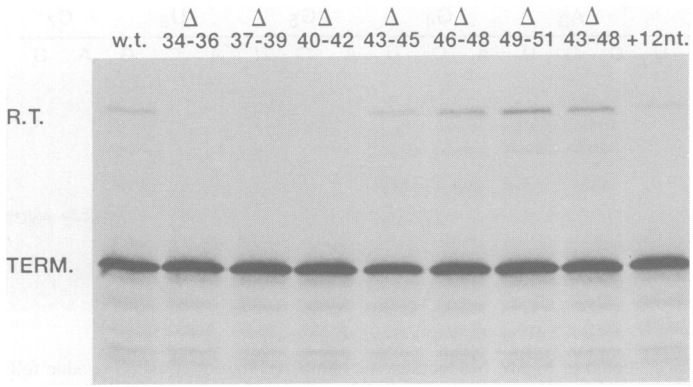

Fig. 5. Effects on read-through of deletions in loop 2. The positions of the deleted nucleotides are shown above the appropriate lane. The lane labelled '+12 nt.' contains an insertion of a random sequence, AUCCUUGUUUCA, 3′ of C45. Termination (TERM.) and read-through (R.T.) products are indicated.

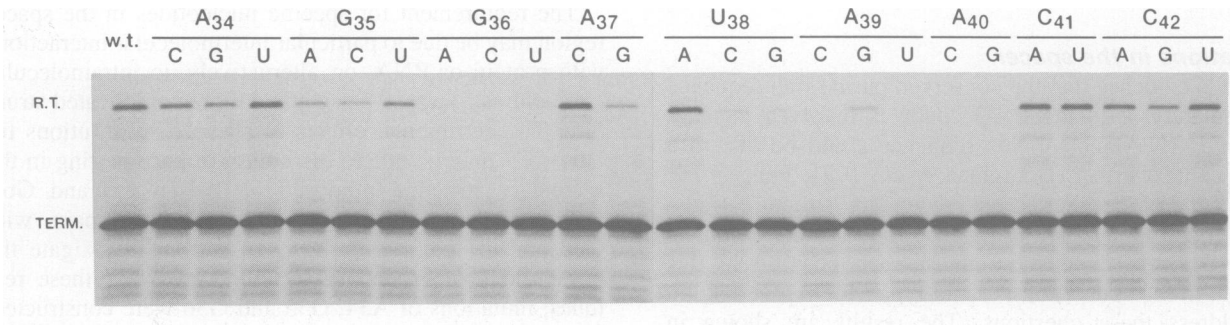

Fig. 6. Effects on read-through of base substitutions in loop 2. The wild-type MuLV sequence is shown on the top line. G35, G36, U38, A39 and A40 are conserved among the read-through retroviruses. The base substitutions tested are shown underneath the appropriate wild-type nucleotide.

G52 as part of stem 2. In this mutant, an additional base pair could form between G17 and C24 making pseudoknot formation unlikely and thereby explaining the lack of read-through observed. These results support the pseudoknot model in which nucleotides A17 and G18 are single-stranded.

Mutations in loop 2

Initial experiments to investigate the importance of loop 2 sequences utilized a set of deletions each lacking three nucleotides. Deletion of AGG34–36, AUA37–39 or ACC40–42 significantly reduced read-through compared with wild-type (Figure 5). Feng *et al.* (1992) reported a similar result for deletion of AGG34–36. The deleterious effect of the three nucleotide deletion cannot be attributed solely to the reduction in the length of loop 2 since other deletions of three nucleotides, CUC43–45, AAA46–48 and GUC49–51, generate as much or more read-through than wild-type (Figure 5). In fact, a larger deletion of six nucleotides (43–48) can be tolerated (Figure 5). Although the size of loop 2 (17–18 nucleotides) is almost constant among the read-through retroviruses, there appears to be some flexibility. A construct containing a 12 nucleotide insertion 3′ to C45, increasing loop 2 to 30 nucleotides, retains ~40% of wild-type read-through activity (Figure 5).

The conclusion from the above results is that some, or all, of the nine nucleotides at positions 34–42 are required for read-through. To identify which particular nucleotides are involved, substitutions were constructed at each of the nine positions. As shown in Figure 6, substitution of G36, U38, A39 or A40 with any of the three other nucleotides reduces read-through significantly except when U38 was replaced by A and read-through was unaffected. It is interesting to note that these four nucleotides are among those conserved in loop 2 among the read-through retroviruses.

There are three highly conserved U residues in loop 2. The results of nucleotide deletions in loop 2 support a role for U38 but argue against the involvement of U44 and U50, which is perhaps surprising considering their conservation. The importance of U44 and U50 was tested directly by base substitutions. As expected, it was found that A, G or C could substitute for U_{50} without affecting read-through (data not shown). A and G had little effect at position U44. When U44 was changed to C, read-through was reduced to ~30% of wild-type (data not shown). This result was unexpected because deletion of CUC43–45 (which includes U44) results in the same level of read-through as wild-type. It should be noted that mutation of U44 to C creates a run of five C residues which could compete with C19-C24 for pairing in stem 2. (However, the mutagenic data argue against equivalent pairing of the wild-type loop 2 sequence. The only difference between the two potential stems is a C-G pair versus a U-G pair.) Of the three conserved U residues in loop 2, only U38 appears critical for the read-through process.

We previously reported that substitution of AAA46–48 with either UAA or UGA, creating an in-frame stop codon in loop 2, increased read-through substantially while substitution with UGG decreased read-through (Wills *et al.*, 1991). The apparent increase in read-through with

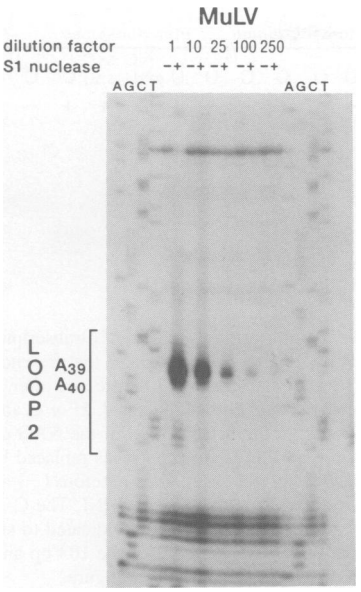

Fig. 7. S1 nuclease sensitivity of the MuLV pseudoknot RNA. The loop 2 region of the wild-type MuLV pseudoknot is bracketed. The sequence of the RNA is shown in outside lanes marked A, G, C and T. The two hypersensitive residues in loop 2 of MuLV, A39 and A40, are shown next to the RNA sequence. The dilution factor of S1 nuclease in each set of reactions is given above the lanes. Each set of reactions contained a control without nuclease (−) and a test reaction (+) exposed to nuclease. The bands that appear in all lanes are strong reverse transcriptase stops.

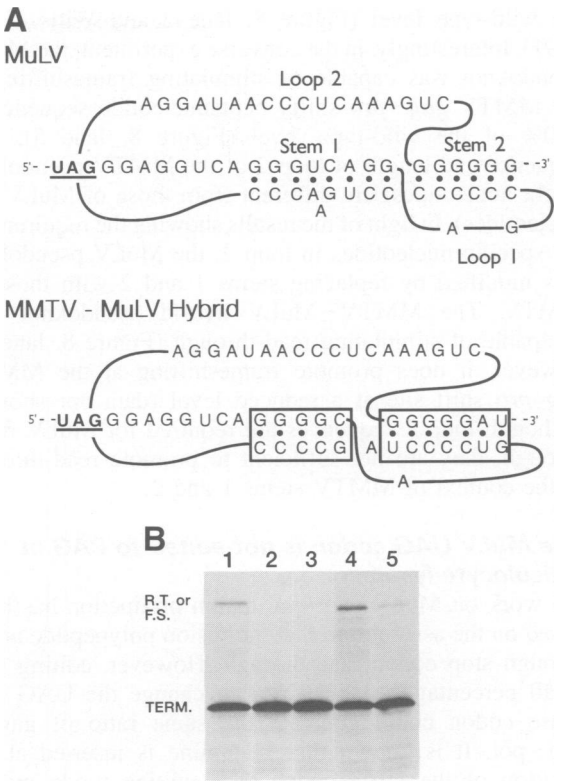

Fig. 8. The effects on read-through of replacing the MuLV pseudoknot with either the MMTV or a hybrid pseudoknot. (**A**) The wild-type MuLV pseudoknot and hybrid MuLV pseudoknot with stems 1 and 2 replaced by those of MMTV are shown. The boxed sequences are the stem regions of the MMTV pseudoknot. (**B**) Read-through assays: lane 1, wild-type MuLV pseudoknot; lane 2, MuLV pseudoknot replaced by MMTV pseudoknot; lane 3, hybrid MuLV−MMTV pseudoknot (see A); lane 4, wild-type MMTV *gag−pro* shift site and pseudoknot; and lane 5, MMTV *gag−pro* shift site and MuLV pseudoknot. The positions are indicated for the termination product (TERM), read-through product (R.T.) for lanes 1–3 and frameshift product (F.S.) for lanes 4 and 5.

either stop codon is likely to have been an over-estimate for the following reason. The distance to the introduced stop codon in loop 2 is 16 codons from the *gag* UAG as opposed to several hundred codons to the end of the transcript in the UGG46–48 construct. Ribosomes reading through the upstream UAG codon and terminating in loop 2 would have less opportunity for premature dissociation, while those that must decode the additional 330 amino acids in the wild-type and UGG46–48 mRNA would have many possibilities for premature termination. This would account for the apparent increase in read-through with an in-frame UAA or UGA in loop 2.

In the mutant with the UGG substitution where read-through is decreased, a stem−loop structure involving UGG and a critical residue, A40, can potentially form. These interactions cannot occur in the wild-type loop 2. Regardless of whether artificially created secondary structures involving loop 2 sequences exist and reduce read-through, the results with the UGG and U44→C mutants show that the identity of a sequence which is dispensable as defined by deletion analysis, can affect the level of read-through.

Sensitivity to S1 nuclease

Structural probing of the wild-type MuLV pseudoknot RNA with the single-strand specific nuclease, S1, revealed hypersensitivity of two residues, A39 and A40, in loop 2 (Figure 7). It has not been established whether both A39 and A40 constitute primary reaction sites with S1. It is possible that the first cleavage of the RNA changes its conformation, unmasking an additional S1-sensitive site. In contrast, mouse mammary tumor virus (MMTV) pseudoknot RNA lacks a predominant S1-cleavage site but instead shows sensitivity of many nucleotides in predicted single-stranded regions (X.Chen, M.Chamorro, S.Lee, I.Tinoco and H.E.Varmus, manuscript in preparation; Wills *et al.*, unpublished data). Interestingly, the two hypersensitive A residues in MuLV loop 2 are among the small number of nucleotides in that region whose identity is important for read-through as shown by substitution analysis above. RNAs generated from constructs of each of the substitutions of A39 or A40 were probed with S1 nuclease. Despite the fact that these constructs, with the possible exception of A39→G, do not promote significant read-through, sensitivity of these positions in loop 2 was retained (unpublished data). The simplest interpretation of these results is that these positions, regardless of the identities of the nucleotides present, are particularly accessible or in an especially reactive conformation.

Hybrid pseudoknot

Chamorro *et al.* (1992) have shown that the stimulator for −1 frameshifting at the *gag−pro* junction of MMTV is a pseudoknot located 3′ to the shifty heptanucleotide sequence. Replacement of the MuLV pseudoknot downstream of the UAG with the MMTV pseudoknot results in read-through but with reduced efficiency of ~10% of

the wild-type level (Figure 8, lane 2 and Wills *et al.*, 1991). Interestingly, in the converse experiment, the MuLV pseudoknot was capable of stimulating frameshifting at the MMTV *gag–pro* shifty heptanucleotide sequence at ~30% of the wild-type level (Figure 8, lane 5). The sequence and length of loop 2 of the MMTV pseudoknot (eight nucleotides) are different from those of MuLV (18 nucleotides). In light of the results showing the requirement for specific nucleotides in loop 2, the MuLV pseudoknot was modified by replacing stems 1 and 2 with those of MMTV. The MMTV–MuLV hybrid pseudoknot was incapable of stimulating read-through (Figure 8, lane 3); however, it does promote frameshifting at the MMTV *gag–pro* shift site at a reduced level (data not shown). Although loop 2 sequences are required for MuLV read-through, they are not sufficient to promote read-through in the context of MMTV stems 1 and 2.

The MuLV UAG codon is not edited to CAG in reticulocyte lysates

All work on MuLV gag–pol fusion production has been based on the assumption that the fusion polypeptide arises through stop codon read-through. However, editing of a small percentage of the RNAs to change the UAG to a sense codon could generate the same ratio of gag to gag–pol. It is known that glutamine is inserted at the position of the UAG codon. The editing model would predict that the UAG terminator is changed to one of the glutamine codons, CAG or CAA. There is no precedent for pseudoknot-dependent editing as would be required in this case, but it is a possibility.

The wild-type transcript was therefore tested for editing of the UAG terminator to CAG. Transcripts were used to program a reticulocyte lysate translation mixture. To address whether putative editing might be translation dependent, i.e. occur on translating ribosomes, the reaction mixtures were separated into ribosome and S-100 supernatant fractions following incubation. RNAs were recovered from both fractions, reverse-transcribed with a MuLV specific primer and amplified by PCR to generate 103 bp products which were digested with *Bst*NI (Figure 9). The transition of U→C creates the recognition sequence for restriction endonuclease *Bst*NI. There was no evidence of editing of the UAG-containing transcripts to CAG by this assay in either the free or ribosome-bound mRNA. The limit of detection was estimated to be 0.1%. PCR products generated from control CAG-containing transcripts were digested completely with *Bst*NI as expected. The PCR products were digested with another restriction endonuclease, *Sty*I, to test for editing of UAG to CAA and none was detected (data not shown). Based on these results, we conclude that the UAG codon is not edited. Therefore, production of the polyprotein requires translational read-through of the UAG codon.

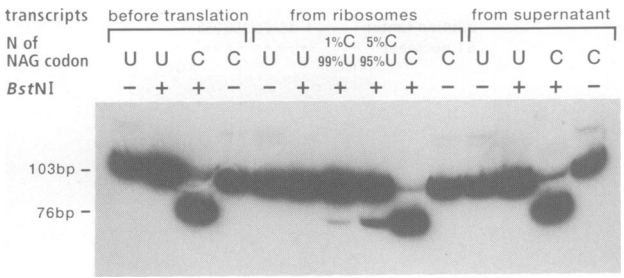

Fig. 9. Assay for editing of the MuLV UAG transcripts. Conversion of the 103 bp PCR product to 76 bp indicates the presence of CAG in the transcript. Transcripts were tested before and after incubation in a reticulocyte translation mixture as indicated. U or C above each lane shows the nucleotide in the first position of the NAG codon; U is found in the wild-type MuLV RNA and was replaced by C in the positive control. PCR products are shown before (–) and after (+) digestion with restriction endonuclease, *Bst*NI. The C- and U-containing PCR products were mixed as indicated to show the sensitivity of the assay. The positions of the 103 bp and 76 bp fragments are indicated at the side of the figure.

Discussion

Amino acid sequencing of the protein product which spans the gag–pol junction of MuLV and FeLV has demonstrated that the UAG termination codon is decoded as glutamine. This precludes splicing of some of the genomic RNA to remove the termination codon as the mechanism for production of the gag–pol fusion polypeptide. However, glutamine would be the expected amino acid if some RNAs (~5%) were edited to change the UAG codon to either CAG or CAA, generating a gag:gag–pol ratio of 20:1. Edited RNAs that were packaged would yield inviable viruses since the normal ratio of gag:gag–pol is crucial (Felsenstein and Goff, 1988; Feng *et al.*, 1989; Jones *et al.*, 1989). It could not be ruled out *a priori* that the virus population tolerates such a small proportion of unproductive particles and so it was necessary to test directly for editing. The results reported above rule out editing of UAG to CAG or CAA in the RNA. Like the retroviruses that utilize frameshifting, MuLV and related viruses use a translational mechanism for generating the gag–pol fusion polypeptide, thus avoiding the disadvantages associated with altering the genomic RNA.

Although it has only been demonstrated for MuLV, pseudoknot-dependent read-through is probably utilized by several other retroviruses noted above for the following reasons: (i) *gag* and *pol* are in the same reading frame separated by a UAG codon, (ii) the sequences 3′ to the UAG codons can potentially form strikingly similar structures, and (iii) not only is the overall pseudoknot structure conserved but many nucleotides in the predicted spacers and single-stranded regions of the pseudoknots are also conserved. The data presented above demonstrate that several of the potential pseudoknot sequences are sufficient to promote read-through. Taken together with previous reports for *in vitro* read-through of MuLV, it is highly likely that read-through in the other retroviruses is also pseudoknot dependent. The consensus sequence of the read-through retroviruses (Figure 2) highlights the conserved nature of several nucleotides in the spacer regions and in loop 2 of the pseudoknots. Mutagenic analysis of these 'conserved' nucleotides in the MuLV pseudoknot showed that many are essential for efficient read-through.

Although MuLV read-through *in vitro* requires the downstream pseudoknot, available *in vivo* data utilizing mutant viruses confirm only the requirement for the first stem (Felsenstein and Goff, 1992). Some disruptive mutations in the proposed stem 2 were viable and all compensatory mutations were inviable. The particular

compensatory mutations tested replaced C-G with A-U base pairs, perhaps altering the stability of the stem in such a way as to disrupt function. In the read-through retroviruses, there is a strong bias for C-G base pairs in stem 2. In the two cases where there is either an A-U (GaLV) or a G-U pair (SNV), stem 2 could potentially consist of a total of 7 bp, 6 of which are C-G. This may reflect the necessity for an extremely stable stem 2, or alternatively, a specific G-rich sequence at the 3′ end of the pseudoknot. A thorough analysis is necessary to reassess the importance of the proposed pseudoknot for read-through *in vivo*.

In many, if not all, instances where stop codon read-through is utilized, sequences immediately 3′ to the stop codon are known to have an effect (Tate and Brown, 1992). Skuzeski *et al.* (1991) identified a six nucleotide sequence, CARYYA, that is required for read-through of the TMV UAG codon (see Zerfass and Beier, 1992). Recent work by Li and Rice (1993) and our unpublished results with T.Tuohy on read-through of UGA in Sindbis virus shows that an even shorter sequence immediately 3′ to the UGA codon promotes read-through. As shown here and in previous studies (Honigman *et al.*, 1991; Feng *et al.*, 1992), nucleotides in the spacer region affect UAG read-through in MuLV. Although the identities of specific nucleotides, particularly U6 and C7, are crucial for read-through, their influence is exerted only in the context of the pseudoknot. In comparing MuLV read-through with TMV and Sindbis, it is interesting that with MuLV, the nucleotides immediately 3′ to the UAG codon are less influential than nucleotides located further downstream and that an elaborate RNA structure is required as a stimulator. In plant luteoviruses, there is also evidence that distant sequences dramatically influence read-through (Miller *et al.*, 1994). Clearly, there are different strategies for achieving read-through.

The length of the spacer between the stop codon and the start of the pseudoknots in the read-through retroviruses, as drawn in Figure 2, is eight nucleotides. In contrast, there is considerable variation in the length of spacers separating shift sites from their respective pseudoknots in retroviruses which utilize frameshifting (ten Dam *et al.*, 1990). The natural length has been shown to be optimal for two cases of −1 frameshifting, IBV (Brierley *et al.*, 1989) and feline immunodeficiency virus (Morikawa and Bishop, 1992), where insertion or deletion of three nucleotides in the spacer drastically lowers frameshifting. Although it is difficult to address the spacing question in MuLV in light of the requirements for specific nucleotides in the spacer, increasing the spacing to 11 nucleotides drastically lowers read-through (our unpublished results). A decrease in the spacing to five nucleotides (by deletion of GGU4–6) also reduces read-through (our unpublished results; Feng *et al.*, 1992) but this effect may be attributed to the removal of two important determinants, G5 and U6. Presumably, the conserved spacing of eight nucleotides is optimal for efficient read-through and may be integral to the mechanism.

Other specific nucleotides, G36, U38, A39 and A40, influence read-through in MuLV. These nucleotides lie in loop 2 of the pseudoknot and are conserved among the read-through retroviruses. Probing of pseudoknot RNA with S1 nuclease showed sensitivity of A39 and A40, suggesting that they are particularly accessible or are in an unusually reactive conformation. This characteristic may offer insights into the structure of the pseudoknot and merits further investigation. These findings are intriguing because of the contrast to the flexibility of sequence in the loops of the pseudoknot that promotes frameshifting in IBV (Brierley *et al.*, 1991).

Although the read-through process requires the downstream pseudoknot, it is not known how the pseudoknot exerts its effect. It has been postulated that pseudoknots cause a pause during translation and pausing has been shown to be important for −1 frameshifting in the yeast L-A virus and IBV, at least, *in vitro* (Tu *et al.*, 1992; Somogyi *et al.*, 1993). The extent of a pseudoknot-induced pause is unlikely to reflect simply the energy required for its unwinding because a stem−loop of greater potential stability causes less pausing (Somogyi *et al.*, 1993). Instead, the difficulty of unwinding a pseudoknot may reflect a feature of its peculiar architecture such as the nucleotides of loops 1 and 2 that cross the helical grooves. It is known that pausing alone is not sufficient to promote frameshifting (Tu *et al.*, 1992; Somogyi *et al.*, 1993). Although it has not been tested, the MuLV pseudoknot may be expected to cause a similar pausing of ribosomes. It is easier to envision how a pause might increase the likelihood of frameshifting where two tRNAs re-pair on the message than to alter the outcome of the competition for the UAG codon between release factor and $tRNA^{Gln}$. It is known that the rate of termination in reticulocyte lysates is markedly slower than elongation, as in other systems (Wolin and Walter, 1988), so a pause near the termination codon would be expected. If the effect of the MuLV pseudoknot is to induce a pause at the termination codon, the question arises as to why a pseudoknot-induced pause is necessary. It is tempting to propose that the pseudoknot alters the termination process directly, or indirectly, through a hypothetical soluble factor, by interaction with rRNA or ribosomal protein(s). The nucleotides in the spacer and loop 2 that affect read-through may be sites of interaction. It is also conceivable that the pseudoknot structure enters the ribosome disrupting its normal conformation, thereby altering the termination process. The latter scenario is theoretically possible since recent data suggest that a complicated RNA structure can actually enter the ribosome (Ringquist *et al.*, 1993).

Many questions remain concerning the pseudoknot-dependent read-through process. What, if any, components of the translational apparatus interact with the RNA structure, what is the molecular nature of the RNA structure, and what is the relative importance of the second stem of the pseudoknot? Chemical and enzymatic probing along with NMR analysis may allow the determination of the RNA structure perhaps providing insights into its mode of action.

Materials and methods

In vitro translations

Complementary oligonucleotides with *Hin*dIII and *Apa*I protruding ends were cloned in-frame into vector pRW201 between the coding regions for CAT and β-galactosidase (*lacZ*) as previously described (Wills *et al.*, 1991). In the MMTV−MuLV hybrid pseudoknot sequence, MMTV stems 1 and 2 could substitute directly for MuLV stems 1 and 2. Plasmid DNAs were purified on CsCl gradients, linearized with *Eco*RV and used

as templates for *in vitro* transcription using T7 RNA polymerase (Promega). Transcripts were translated *in vitro* in a reticulocyte lysate system (Promega) and the products separated on 15% SDS–polyacrylamide gels. Dried gels were exposed for several hours or overnight. Autoradiograms were scanned on a Molecular Dynamics densitometer. Corrections were made for the differential amounts of [^{35}S]methionine in the two products. The read-through levels are expressed as a percentage of the wild-type level as performed previously (Wills *et al.*, 1991).

RNA editing analysis

RNA transcripts were generated from clones containing the wild-type MuLV sequence (UAG) and the positive control sequence (CAG). The transcripts were incubated in a reticulocyte translation mixture at 30°C for 1 h. Reaction mixtures were fractionated by centrifugation at 73 000 r.p.m. for 2 h at 4°C in a Beckman TLA 100.2 rotor. RNA was recovered from the supernatant and ribosome pellets by extraction with phenol/chloroform and precipitation with ethanol. cDNAs were made from the recovered RNAs, using a primer which annealed 3′ to the (U/C)AG, and used as templates for PCR in which one of the primers was 5′ end-labelled with ^{32}P. The 103 bp PCR products were gel purified, digested with restriction endonucleases *Bst*NI [recognition sequence: CC(A/T)GG] or *Sty*I [recognition sequence: CC(A/T)(A/T)GG] obtained from New England Biolabs, and analyzed on 12% native polyacrylamide gels. Conversion of the 103 bp fragment to 76 bp is indicative of CAG (*Bst*NI) or CAA (*Sty*I) in the RNA.

S1 nuclease analysis

In vitro transcripts were resuspended in 100 µl 5 mM MES pH 6.3, 5 mM $MgCl_2$, 60 mM NaCl (Wyatt *et al.*, 1990). A 50 µl sample was removed and extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and precipitated with 5 µg glycogen, 5 µl of 3 M NaOAc and 2 vol of ethanol. Precipitated RNAs were washed with 70% ethanol and lyophilized. RNAs treated with S1 nuclease (280 U/µl, Pharmacia) were processed as above after incubation for 10 min at room temperature with the enzyme (M.Chamorro, personal communication). Primer extensions of the RNAs were done using an oligonucleotide complementary to 18 nucleotides of *lacZ* and MuLV reverse transcriptase (Life Sciences Inc.). Products were separated on a 6% denaturing polyacrylamide gel (acrylamide:bisacrylamide; 19:1). RNA sequences were determined using the *lacZ* primer, MuLV reverse transcriptase and dideoxynucleotides.

Acknowledgements

We thank Dr Judith Levin for providing clones, Dr Mark Sandbaken for useful discussions, Robin Johnston for technical assistance and the members of the Gesteland lab for continual encouragement. The idea for probing the MuLV pseudoknot RNA with S1 nuclease arose from discussions with Mario Chamorro. This work was supported by the Howard Hughes Medical Institute and National Institutes of Health grant # 1-RO1-GM41852.

References

Atkins,J.F., Weiss,R.B. and Gesteland,R.F. (1990) *Cell,* **62,** 413–423.
Berry,M.J., Banu,L., Harney,J.W. and Larsen,P.R. (1993) *EMBO J.,* **12,** 3315–3322.
Böck,A., Forchhammer,K., Heider,J. and Baron,C. (1991) *Trends Biochem. Sci.,* **16,** 463–467.
Brierley,I., Digard,P. and Inglis,S.C. (1989) *Cell,* **57,** 537–547.
Brierley,I., Rolley,N.J., Jenner,A.J. and Inglis,S.C. (1991) *J. Mol. Biol.,* **220,** 889–902.
Capone,J.P., Sedivy,J.M., Sharp,P.A. and RajBhandary,U.L. (1986) *Mol. Cell. Biol.,* **6,** 3059–3067.
Chamorro,M., Parkin,N. and Varmus,H.E. (1992) *Proc. Natl Acad. Sci. USA,* **89,** 713–717.
Delassus,S., Sonigo,P. and Wain-Hobson,S. (1989) *Virology,* **173,** 205–213.
Felsenstein,K.M. and Goff,S.P. (1988) *J. Virol.,* **62,** 2179–2182.
Felsenstein,K.M. and Goff,S.P. (1992) *J. Virol.,* **66,** 6601–6608.
Feng,Y.-X., Levin,J.G., Hatfield,D.L., Schaefer,T.S., Gorelick,R.J. and Rein,A. (1989) *J. Virol.,* **63,** 2870–2873.
Feng,Y.-X., Yuan,H., Rein,A. and Levin,J.G. (1992) *J. Virol.,* **66,** 5127–5132.
Hatfield,D.L., Levin,J.G., Rein,A. and Oroszlan,S. (1992) *Adv. Virus Res.,* **41,** 193–239.
Honigman,A., Wolf,D., Yaish,S., Falk,H. and Panet,A. (1991) *Virology,* **183,** 313–319.
Jones,D.S., Nemoto,F., Kuchino,Y., Masuda,M., Yoshikura,H. and Nishimura,S. (1989) *Nucleic Acids Res.,* **17,** 5933–5945.
Kato,S., Matsuo,K., Nishimura,N., Takahashi,N. and Takano,T. (1987) *Jap. J. Genet.,* **62,** 127–137.
Li,G. and Rice,C.M. (1993) *J. Virol.,* **67,** 5062–5067.
Miller,W.A., Dinesh-Kumar,S.P. and Paul,C.P. (1994) *CRC Crit. Rev. Plant Sci.,* in press.
Morikawa,S. and Bishop,D.H.L. (1992) *Virology,* **186,** 389–397.
Perryman,S.M., McAtee,F.J. and Portis,J.L. (1991) *Nucleic Acids Res.,* **19,** 1707.
Philipson,L., Andersson,P., Olshevsky,U., Weinberg,R., Baltimore,D. and Gesteland,R. (1978) *Cell,* **13,** 189–199.
Pleij,C.W.A., Rietveld,K. and Bosch,L. (1985) *Nucleic Acids Res.,* **13,** 1717–1731.
Reddy,E.P., Smith,M.J. and Aaronson,S.A. (1981) *Science,* **214,** 445–450.
Ringquist,S., MacDonald,M., Gibson,T. and Gold,L. (1993) *Biochemistry,* **32,** 10254–10262.
Skuzeski,J.M., Nichols,L.M., Gesteland,R.F. and Atkins,J.F. (1991) *J. Mol. Biol.,* **218,** 365–373.
Somogyi,P., Jenner,A.J., Brierley,I. and Inglis,S.C. (1993) *Mol. Cell. Biol.,* **13,** 6931–6940.
Tate,W.P. and Brown,C.M. (1992) *Biochemistry,* **31,** 2443–2450.
ten Dam,E.B., Pleij,C.W.A. and Bosch,L. (1990) *Virus Genes,* **4,** 121–136.
Tu,C., Tzeng,T.-H. and Bruenn,J.A. (1992) *Proc. Natl Acad. Sci. USA,* **89,** 8636–8640.
Van Beveren,C., van Straaten,F., Galleshaw,J.A. and Verma,I.M. (1981) *Cell,* **27,** 97–108.
Weaver,T.A., Talbot,K.J. and Panganiban,A.T. (1990) *J. Virol.,* **64,** 2642–2652.
Wills,N.M., Gesteland,R.F. and Atkins,J.F. (1991) *Proc. Natl Acad. Sci. USA,* **88,** 6991–6995.
Wolin,S.L. and Walter,P. (1988) *EMBO J.,* **7,** 3559–3569.
Wyatt,J.R., Puglisi,J.D. and Tinoco,I.,Jr (1990) *J. Mol. Biol.,* **214,** 455–470.
Yoshinaka,Y., Katoh,I., Copeland,T.D. and Oroszlan,S. (1985a) *Proc. Natl Acad. Sci. USA,* **82,** 1618–1622.
Yoshinaka,Y., Katoh,I., Copeland,T.D. and Oroszlan,S. (1985b) *J. Virol.,* **55,** 870–873.
Zerfass,K. and Beier,H. (1992) *Nucleic Acids Res.,* **20,** 5911–5918.

Received on May 6, 1994; revised on June 7, 1994