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.

Cell, Vol. 57, 537–547, May 19, 1989, Copyright © 1989 by Cell Press

Characterization of an Efficient Coronavirus Ribosomal Frameshifting Signal: Requirement for an RNA Pseudoknot

Ian Brierley, Paul Digard, and Stephen C. Inglis
Division of Virology
Department of Pathology
University of Cambridge
Cambridge CB2 1QP
England

Summary

The genomic RNA of the coronavirus IBV contains an efficient ribosomal frameshifting signal at the junction of two overlapping open reading frames. We have defined by deletion analysis an 86 nucleotide sequence encompassing the overlap region which is sufficient to allow frameshifting in a heterologous context. The upstream boundary of the signal consists of the sequence UUUAAAC, which is the likely site of ribosomal slippage. We show by creation of complementary nucleotide changes that the RNA downstream of this "slippery" sequence folds into a tertiary structure termed a pseudoknot, the formation of which is essential for efficient frameshifting.

Introduction

Ribosomal frameshifting is a recently described mechanism of translational regulation in which a directed change of translational reading frame allows the synthesis of a single protein from two (or more) overlapping genes (see Roth, 1981, Craigen and Caskey, 1987, for reviews). So far, almost all the examples of this kind of control in higher eukaryotes have come from retroviruses, where frameshifting appears to be a mechanism for the regulation of expression of the viral RNA-dependent DNA polymerase. One termination codon in Rous sarcoma virus (RSV) (Jacks and Varmus, 1985) and human immunodeficiency virus type 1 (HIV-1) (Jacks et al., 1988a), and two in mouse mammary tumor virus (MMTV) (Moore et al., 1987; Jacks et al., 1987) are suppressed by (−1) ribosomal frameshifts into alternative overlapping open reading frames (ORFs), generating the *gag/pol* polyproteins from which the viral polymerases are subsequently derived. In addition, frameshifting appears to be necessary for the expression of the reverse transcriptase enzymes of a number of retrotransposons, such as yeast Ty1 (Mellor et al., 1985; Wilson et al., 1986; Clare et al., 1988) and Ty912 (Clare and Farabaugh, 1985) and the Drosophila 17.6 (Saigo et al., 1984) and gypsy (Marlor et al., 1986) elements.

Recently, however, we described the first non-retroviral, higher eukaryotic example of the phenomenon, in the avian coronavirus infectious bronchitis virus (IBV) (Brierley et al., 1987). IBV, the type species of the Coronaviridae (Siddell et al., 1983), has a large, single-stranded, continuous RNA genome of positive polarity (Schochetman et al., 1977; Stern and Kennedy, 1980). Three-quarters of the genomic coding capacity is contained in two large ORFs, F1 and F2, which potentially encode 440 kd and 300 kd polypeptides respectively (Boursnell et al., 1987). The ORFs of F1 and F2 overlap by 42 nucleotides, with F2 in the −1 frame with respect to F1 (see Figure 1). Using in vitro transcription and translation, we demonstrated that F2 is produced as a fusion protein with F1 as a result of a highly efficient (25%–30%) (−1) ribosomal frameshift that suppresses the F1 termination codon (Brierley et al., 1987). The role of this translational mechanism in IBV replication is so far unknown, but since the F1 and F2 ORF are assumed to encode, at least in part, components of the virus-specific RNA-dependent polymerase, and since the predicted amino acid sequence of F2 contains regions in common with other positive strand virus replicases (Boursnell et al., 1987; Hodgman, 1988), it may be that the overall strategy is similar to that understood for retroviruses.

There is considerable interest in the precise mechanism by which ribosomal frameshifting operates, and work from several groups has shown that the specificity for the event resides in the nucleotide sequence of the RNA around the site at which frameshifting occurs, since it has proved possible to induce frameshifting in a heterologous context by inserting cloned DNA corresponding to the frameshift site into unrelated genes (Brierley et al., 1987; Jacks et al., 1988b). For RSV and HIV, the site at which frameshifting occurs has been identified by a combination of site-directed mutagenesis and amino acid sequence analysis (Jacks et al., 1988a, 1988b). These authors, from a comparative analysis of a large number of retroviruses thought to utilize frameshifting as a means of controlling gene expression, have suggested that certain heptanucleotide RNA sequences, initiating with two homopolymeric triplets, can allow tRNA slippage during translation, leading to a (−1) frameshift. In addition to these "slippery" sequences, potential RNA stem–loop structures located downstream of most retroviral shift sites have been proposed to make contributions to the frameshifting process (Rice et al., 1985; Moore et al., 1987; Jacks et al., 1987, 1988a), and indeed, the presence of a stem–loop downstream of the RSV frameshift site has been shown to be essential (Jacks et al., 1988b). How the stem–loop influences frameshifting is not known, but it has been suggested (Jacks et al., 1988b) that ribosomes may slow or stall at the stem–loop, increasing the likelihood of a tRNA slippage event.

In this paper we describe analysis of the RNA sequences responsible for frameshifting in IBV, and define the minimal amount of information from the F1/F2 overlap that is sufficient to induce frameshifting at high efficiency. The upstream limit of this essential region has the characteristics of a "slippery" sequence as defined previously (Jacks et al., 1988b), and mutational analysis of this sequence is entirely consistent with frameshifting occurring at this site. The remainder of the essential region consists of about 80 nucleotides immediately downstream of the

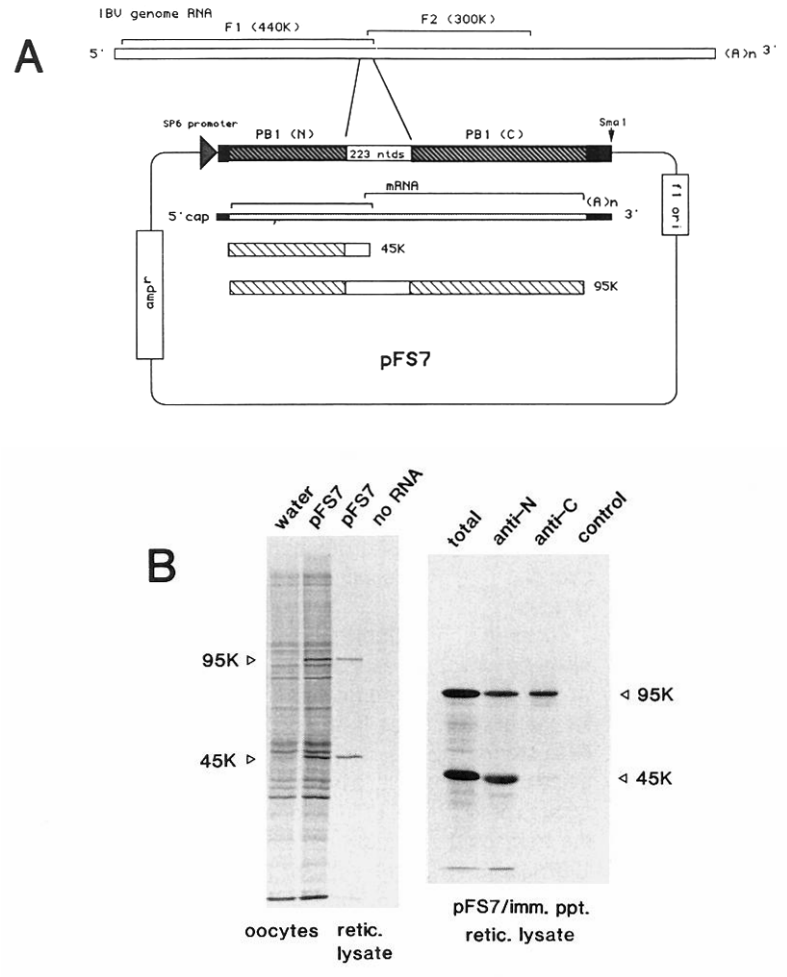

Figure 1. Ribosomal Frameshifting In Vitro and In Vivo

(A) Diagram of plasmid pFS7 showing the predicted sizes of protein products that would be expected following ribosomal frameshifting within the F1/F2 overlap during translation of a mRNA derived from SmaI-digested pFS7 template. (B) Left-hand panel: reticulocyte lysate and Xenopus oocyte translation products synthesized in response to RNA transcribed from SmaI-digested pFS7. RNA was translated and products were labeled with [^{35}S]methionine as described in Experimental Procedures. Polypeptides were separated on a 10% SDS–polyacrylamide gel and detected by autoradiography. Right-hand panel: reticulocyte lysate translation products directed by pFS7-derived mRNA were analyzed directly (total) or immunoprecipitated with antiserum raised against the N- or C-terminus of the PB1 protein, or with a control antiserum (a polyclonal antiserum to the influenza PA protein) as indicated. Polypeptides were labeled with [^{35}S]methionine, separated on 10% SDS–polyacrylamide gels, and detected by autoradiography.

"slippery" sequence, and we present evidence that efficient frameshifting depends on the formation, by these sequences, of a tertiary RNA structure in the form of a "pseudoknot" (Studnicka et al., 1978; Pleij et al., 1985).

Results

Ribosomal Frameshifting In Vitro and In Vivo

A ribosomal frameshifting mechanism for the expression of the IBV F2 ORF was first suggested by Boursnell and colleagues (1987). As no products encoded by F1 and F2 have been detected in IBV-infected cells, we tested this hypothesis by in vitro transcription and translation of a cDNA containing the F1/F2 overlap region cloned within a marker gene (PB2 of influenza virus A/PR8/34) under the control of the bacteriophage SP6 promoter (Brierley et al., 1987). Translation of synthetic mRNA derived from this plasmid (pFS1) in vitro yielded products whose size and pattern of immunoprecipitation (with specific antisera raised against defined portions of the marker protein) were entirely consistent with an efficient (−1) ribosomal frameshifting event at the end of the F1 ORF, resulting in the production of an F1-F2 *trans*-frame protein (Brierley et al., 1987). These earlier studies, however, did not test if the frameshift signal was functional in vivo. To address this question, we wished to test the capacity of our synthetic mRNA to signal frameshifting in Xenopus oocytes after microinjection. However, since the influenza PB2 protein is relatively unstable in Xenopus oocytes, we constructed an alternative recombinant plasmid, pFS7, in which the IBV F1/F2 overlap was cloned within the PB1 gene of influenza virus, which in turn was flanked by the 5′ and 3′ noncoding regions of the Xenopus β-globin gene, downstream of an SP6 promoter (Krieg and Melton, 1984) (Figure 1A). Linearization of this plasmid with SmaI, followed by transcription, resulted in the production of a capped and polyadenylated 2.8 kb mRNA, designed such that on translation, ribosomes terminating at the F1 ORF stop codon would produce a 45 kd product (the "stopped" product) and those that frameshifted, a 95 kd product (see Figure 1A). Since the region of IBV cDNA included in pFS7 was considerably shorter than that in pFS1 (230 bp, as opposed to 497 bp,), we first tested that the pFS7-derived mRNA contained a functional frameshift signal by translating it in the reticulocyte lysate system. As can be seen in Figure 1B, frameshifting occurred at wild-type (25%) efficiency. The predicted 45 kd "stopped" product migrates as a somewhat larger species (about 50 kd) on SDS–poly-

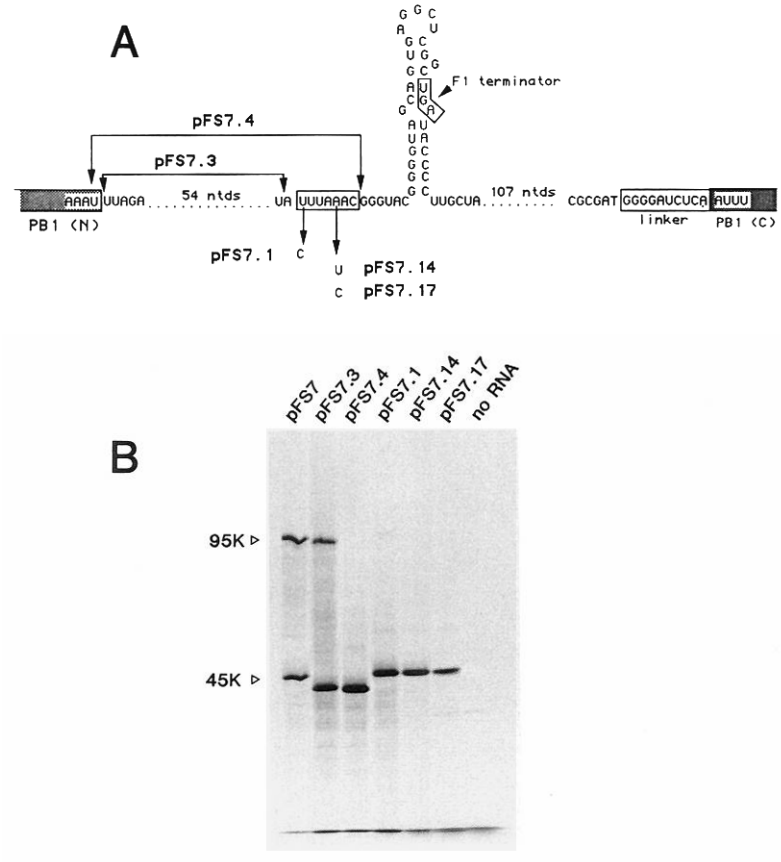

Figure 2. Upstream Components of the IBV Frameshifting Signal

(A) The position of specific deletions and point mutations created within the F1/F2 junction region of plasmid pFS7 by site-directed mutagenesis. The diagram shows part of the sequence of the RNA around the frameshift site, including the suspected "slippery" sequence UUUAAAC (boxed) and the predicted stem–loop structure downstream. Shaded areas represent regions of the influenza PB1 gene, and unshaded areas, IBV-derived sequences, except for a linker sequence (boxed) that was derived from phage m13 during the cloning procedure (see Experimental Procedures). (B) Reticulocyte lysate translation products synthesized in response to mRNAs derived from SmaI-digested pFS7 or mutant templates. Polypeptides were labeled with [^{35}S]methionine, separated on a 10% SDS–polyacrylamide gel, and detected by autoradiography.

acrylamide gels, probably a consequence of the highly basic nature of the PB1 protein. As expected, antiserum as1, raised against the N-terminus of PB1, immunoprecipitated both the 45 kd and 95 kd products, indicating common N-termini, and the C-terminal antiserum, as4, only the 95 kd product. When the same mRNA was microinjected into Xenopus oocytes, frameshifting was also observed, with the stopped and frameshifted products clearly visible against the background of labeled oocyte proteins (see Figure 1B). The efficiency of frameshifting in vivo was similar to that seen in the reticulocyte lysate.

Defining the Upstream Boundary of the Frameshifting Signal

Retroviral frameshifting appears to involve at least two components, a "slippery" sequence, which is the site of tRNA slippage, positioned upstream of a potential stem–loop structure in the mRNA (Jacks et al., 1988b). Analysis of the IBV F1/F2 overlap reveals a similar sequence motif, UUUAAAC, some 6 bp upstream of a potential stem–loop (see Figure 2 and Brierley et al., 1987). Based on the original hypothesis of Jacks and Varmus (1985) on the mechanism of frameshifting in RSV, we suspected that the site of frameshifting in IBV would be the UUUA sequence, with a tRNA slipping from the UUA codon back to the UUU codon in the −1 frame during the frameshifting process. This possibility was tested by site-directed mutagenesis using pFS7, which contains the bacteriophage f1 intergenic region in a nonessential region of the vector (see Experimental Procedures), and which can thus be converted to single-stranded DNA (Dotto et al., 1981). We first created a deletion from within the upstream PB1 sequences up to, but not including, the UUUAAAC motif of the IBV frameshift region (pFS7.3), and another from the same point to just beyond the UUUAAAC motif (pFS7.4) (Figure 2). In each case the deletions were designed to preserve the upstream and downstream open reading frames present in the original plasmid. Synthetic mRNA was transcribed from these plasmids and tested for frameshifting by in vitro translation (Figure 2). The results confirmed the suspicion that the UUUAAAC motif was important, since frameshifting occurred with wild-type efficiency in pFS7.3 but was totally abolished with pFS7.4.

To investigate in more detail the role of this sequence, we made single base substitutions at certain positions within this region. In the first mutagenic construct, pFS7.1, UUUA was changed to CUUA, on the basis that a UUA to CUU tRNA^{Leu} slippage event ought to be unfavorable. As can be seen in Figure 2, with this mutant, frameshifting in the reticulocyte lysate system was reduced approximately 10-fold, supporting the idea that a tRNA slips at this site. However, as frameshifting was not completely abolished, we tested whether an additional site of slippage was involved, at the adjacent AAC codon. To avoid

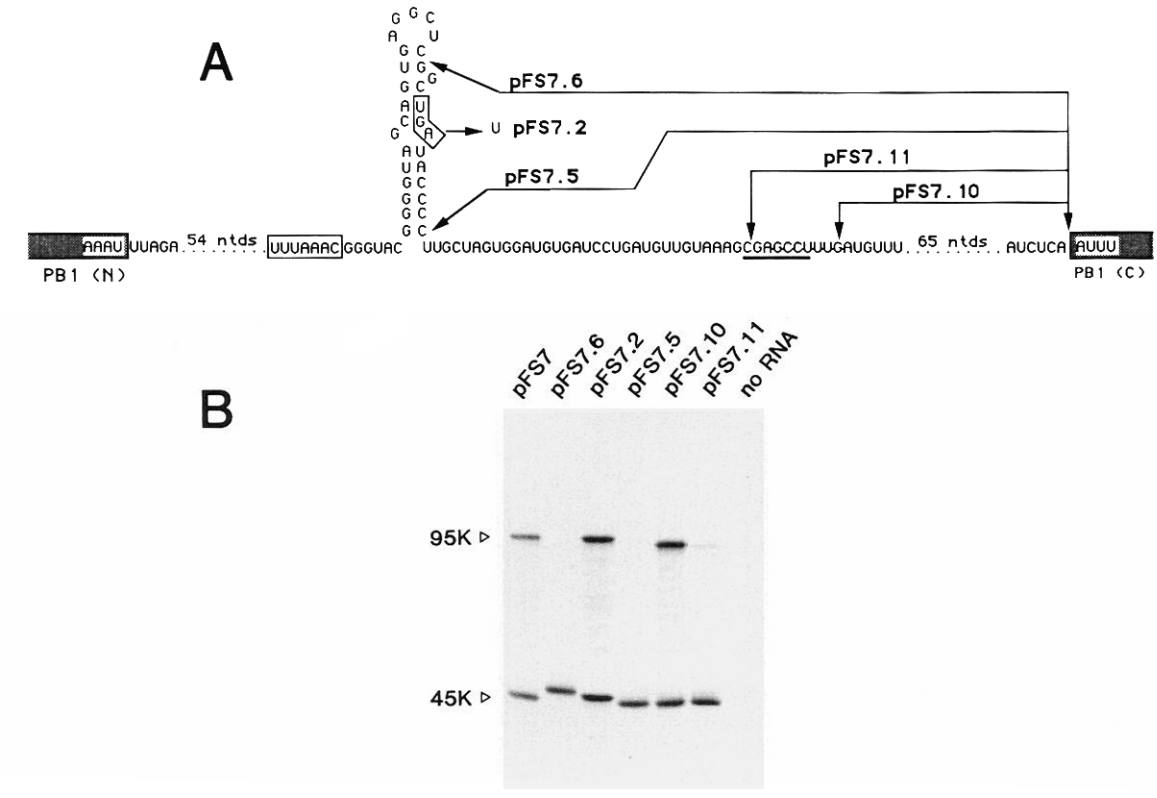

Figure 3. Downstream Components of the IBV Frameshifting Signal

(A) The position of specific deletions and a point mutation created within the F1/F2 region of plasmid pFS7 by site-directed mutagenesis. The diagram shows part of the sequence of the RNA from the overlap region. The underlined region highlights nucleotides with the potential to base pair with complementary nucleotides in the loop of the stem–loop structure predicted to form upstream.

(B) Reticulocyte lysate translation products synthesized in response to mRNAs derived from SmaI-digested pFS7 or mutant templates. Polypeptides were labeled and analyzed as described in the legend to Figure 2.

perturbing the UUUA slip site, we altered the second nucleotide of the AAC codon to either U (pFS7.14) or C (pFS7.17) and tested for frameshifting (Figure 2). Frameshifting was not observed with pFS7.14 (UUUAUAC), and was greatly reduced (to about 0.5%) with pFS7.17 (UUUACAC). The great reduction in frameshifting efficiency arising from these changes suggests that the tRNAs specified by the mutant sequences, i.e., $tRNA^{Tyr}$ (for pFS7.14) and $tRNA^{His}$ (for pFS7.17) are unable to shift into the −1 frame. This is not unreasonable, since such a shift would require the $tRNA^{Tyr}$ anticodon, AUG, to base pair with AUA in the mutant mRNA, and the $tRNA^{His}$ anticodon, GUG, to pair with ACA (the anticodon sequences shown are based on standard Watson–Crick base pairing). In the pFS7.1 mutant, where low-level frameshifting was seen, the $tRNA^{Leu}$ anticodon, AAU, could form one base pair with the mutant codon CUU; this may be sufficient to allow a low efficiency slip. Recently, Jacks and colleagues (1988b) have proposed that in RSV, a double tRNA-slippage event occurs at the frameshift site, in which two adjacent ribosome-bound tRNAs slip into the −1 frame at the sequence AAAUUUA. Our results with IBV are consistent with this simultaneous slippage model, with the adjacent $tRNA^{Leu}$ and $tRNA^{Asn}$ slipping back by one nucleotide in the 5′ direction at the UUUAAAC site in IBV.

Requirement for Downstream Sequences

We next created a series of deletions downstream of the "slippery" sequence to define the 3′ limit of the frameshifting signal (Figure 3). Once again deletions were made by oligonucleotide-directed mutagenesis and were designed to preserve the original open reading frame. We expected on the basis of previous work that this analysis would reveal a requirement for the nucleotides predicted to form a stem–loop structure just downstream of the slippery sequence, and indeed, pFS7.6, in which one side of the predicted stem–loop was removed, displayed virtually no frameshifting. This reduction was not simply due to the removal of the F1 stop codon in pFS7.6, since alteration of the stop codon from UGA to UGU by mutagenesis in a separate construct, pFS7.2, resulted in a mRNA with slightly increased frameshifting efficiency (Figure 3). In this case a new termination codon is used which is located some 36 nucleotides downstream. Thus it seems that precise positioning of a terminator is not a prerequisite for frameshifting, in agreement with the situation for RSV (Jacks et al., 1988b). Surprisingly, pFS7.5, which pre-

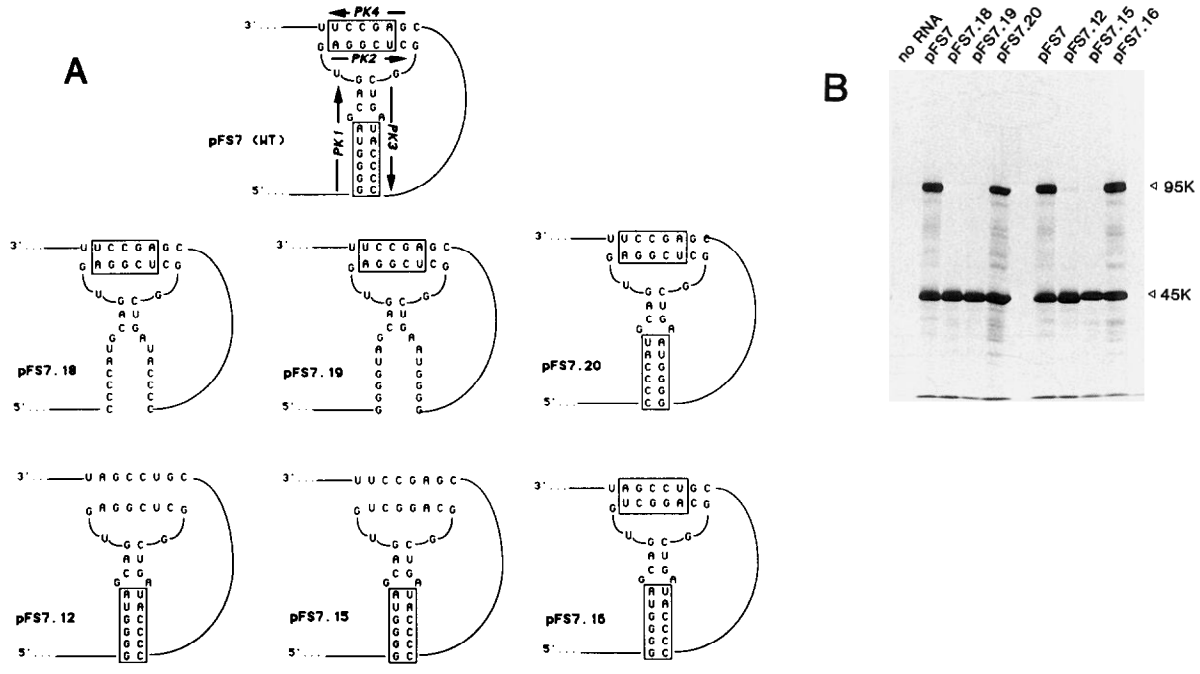

Figure 4. A Tertiary RNA Structure Forms in the IBV Frameshift Region

(A) Diagram showing the predicted base pairing interactions within the RNA at the frameshift site in the wild-type (pFS7/WT) and in six mutant constructs prepared by site-directed mutagenesis. The mutations in constructs pFS7.18, pFS7.19, and pFS7.20 were in the first stem (PK1/PK3), and in constructs pFS7.12, pFS7.15, and pFS7.16, the putative second stem (PK2/PK4) was altered.

(B) Reticulocyte lysate translation products synthesized in response to mRNAs derived from SmaI-digested pFS7 or mutant templates. Polypeptides were labeled and analyzed as described in the legend to Figure 2.

serves the potential for stem–loop formation, also showed greatly reduced frameshifting, although in this case a residual amount, about 2%–3%, was still observed. In light of this observation, we considered the possibility that an additional interaction may occur between the stem–loop structure and nucleotides downstream to create a tertiary structure in the mRNA. Structures of this kind have been proposed to occur in a number of positive-stranded viral RNAs (see Pleij et al., 1985; Clarke et al., 1987) and have been termed "pseudoknots." The formation of a pseudoknot necessitates the base pairing of nucleotides in a hairpin loop with nucleotides downstream, generating an extended quasi-continuous double helix. Close examination of the IBV F1/F2 mRNA sequence revealed a stretch of 7 nucleotides (underlined) about 30 nucleotides downstream of the predicted stem–loop which could potentially base pair with a stretch of nucleotides in the single-strand region of the hairpin loop. As can be seen in Figure 3, frameshifting on pFS7.10-derived mRNA, which retains this sequence, occurred with wild-type efficiency. However, when most of this sequence was deleted, in construct pFS7.11, the frameshifting efficiency was dramatically reduced, which indicates that the nucleotides from this region are crucial for efficient frameshifting.

The Frameshifting Signal Contains a Tertiary RNA Structure

The formation of a pseudoknotted RNA structure downstream of the UUUAAAC sequence would require two discrete base pairing interactions within the mRNA (as shown in Figure 4), one which would form the original stem–loop (between PK1 and PK3), and a second between the loop of this structure (PK2) and the AGCCU-containing region downstream (PK4). As a first step, we sought direct evidence for the PK1–PK3 interaction, to confirm the existence of the first stem–loop, by creating complementary mutagenic changes within the PK1 and PK3 regions (see Figure 4). In pFS7.18, we sought to destabilize the stem by replacing six nucleotides of the PK1 sequence at the base of the stem (GGGGUA) with their complementary nucleotides. In pFS7.19, the six nucleotides of the PK3 sequence at the base of the stem (CCCCAU) were replaced by complementary nucleotides. Finally, in pFS7.20, both mutagenic changes were made, to create a double mutant in which the changes whould recreate base pairing and restabilize the stem. In both the single mutants (pFS7.18 and 7.19), frameshifting was abolished, yet the frameshifting efficiency returned to near wild-type levels in the double mutant, pseudo-wild-type construct, pFS7.20 (Figure 4). This result provides strong support for the PK1–PK3 interaction, and demonstrates that the stem–loop plays an important role in the frameshifting process.

We then went on to investigate the proposed PK2–PK4 interaction in a similar manner. In this case, we mutated the relevant regions by reversing a block of five nucleotides within each region (see Figure 4). Thus in pFS7.15, the PK2 sequence was changed from AGGCU to UCGGA,

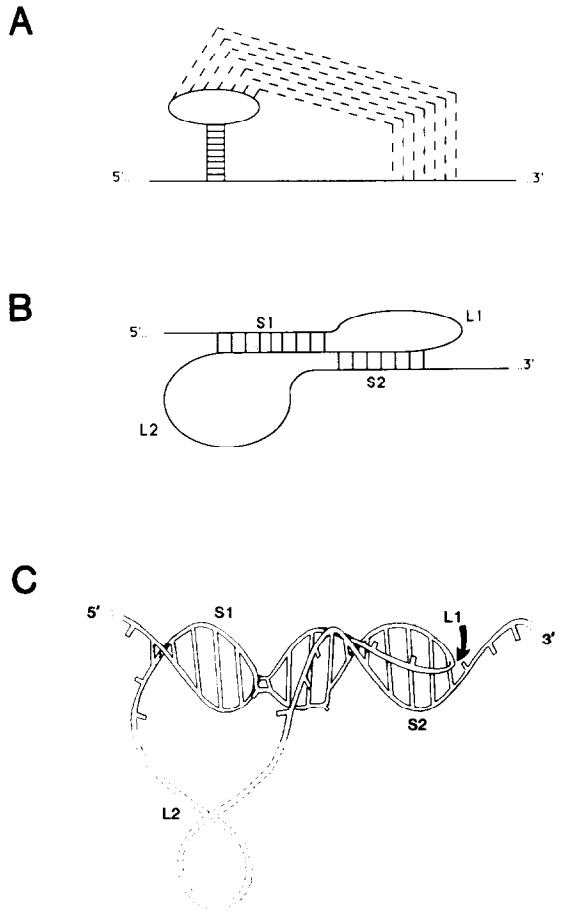

Figure 5. Proposed Folding of the RNA at the IBV Frameshift Region: Formation of an RNA Pseudoknot

Base pairing between nucleotides in the loop of the stem–loop structure and a region downstream (A) results in the formation of an extended double helix, shown schematically (B). The double helical regions S1 and S2 are connected by single-stranded loops L1 and L2. In this structure, S2 is stacked upon S1 such that a right-handed, quasi-continuous double-helix of 16 bp is formed. An artist's impression of the three-dimensional organization of this structure is shown (C), assuming that 9 bp form in S1, 7 bp form in S2, and one turn of the helix contains 11 bp (Arnott et al., 1972). In the resulting pseudoknot, L1 (2 nucleotides in length) crosses the deep groove and L2 (32 nucleotides in length) the shallow groove. The "fold" program of Jacobson et al. (1984) did not predict any significant RNA secondary structures within L2. These diagrams are based on those presented by Rietveld et al. (1983) and Pleij et al. (1985).

and in pFS7.12, the PK4 sequence was changed from AGCCU to UCCGA. In both mutated constructs, where the potential for base pairing is dramatically reduced, the frameshifting efficiency was lowered some 10-fold to about 2%. However, as was the case with the PK1-PK3 interaction, the creation of the double mutant, pFS7.16, in which the changes are complementary, restored the efficiency of frameshifting to approximately wild-type level. In these experiments, it was consistently observed that the "stopped" product from pFS7.15 and pFS7.16 had a slightly lower gel mobility than that produced by the wild-type and pFS7.12 constructs. This altered gel mobility is most likely due to the two amino acid substitutions (serine and aspartic acid replacing arginine and leucine) arising from mutation of the PK2 sequence in these plasmids. These experiments provide strong evidence that a tertiary RNA structure, a pseudoknot, forms at the frameshift site. A model for such a structure is shown in Figure 5.

The Slippery Sequence and Pseudoknot Are Precisely Positioned

Jacks and colleagues (1988b) have proposed that RNA secondary structure downstream of the "slippery" site may slow or stall ribosomes during translation, such that the ribosomal peptidyl (P) and acceptor (A) sites remain in the vicinity of the slippery sequence slightly longer, promoting the slippage event. A prediction of such a model is that the correct spacing distance must be maintained between the slippery sequence and the RNA structure concerned. We tested this by constructing two mutants in which the spacing was increased (pFS7.13) or decreased (pFS7.22) by three nucleotides (Figure 6). In pFS7.13, in which a tyrosine codon (UAC) was inserted between the slippery sequence and pseudoknot, frameshifting was abolished. In pFS7.22, deletion of the glycine codon (GGG) greatly reduced frameshifting (to about 2%). Thus alteration of the relative position of the UUUAAAC sequence by as little as three nucleotides with respect to the pseudoknot causes a dramatic reduction in frameshifting efficiency.

Discussion

The genomic RNA of the coronavirus IBV contains a sequence at the junction of the F1 and F2 ORF that can direct ribosomal frameshifting both in vitro and in vivo in Xenopus oocytes. We have defined, by deletion analysis, a stretch of approximately 86 nucleotides within this region that contains all the necessary information for high efficiency frameshifting in a heterologous context. This minimal frameshift signal includes, at its 5′ end, the heptameric sequence UUUAAAC, which appears to be analagous to the "slippery sequences" described by Jacks et al. (1988a, 1988b) required for frameshifting in retroviral systems. In addition, frameshifting is dependent upon the formation of a tertiary mRNA structure downstream of the slippery sequence. This represents direct evidence that tertiary folding in RNA can influence its translation and could have implications for a much wider range of biological processes involving RNA. The frameshifting system thus provides an excellent opportunity to study in more detail this novel kind of RNA structure.

The Site of Frameshifting in Coronavirus IBV

Our results suggest strongly that frameshifting occurs at the UUUAAAC sequence. A deletion that removed this region abolished frameshifting, and point mutations within the region were severely inhibitory. This sequence corresponds in organization to the "slippery" sequences present in the gene overlaps of demonstrated and suspected retroviral frameshift sites (Jacks and Varmus, 1988b). In the three retroviral overlaps where the site of slippage has been established by protein sequencing, the shift occurs

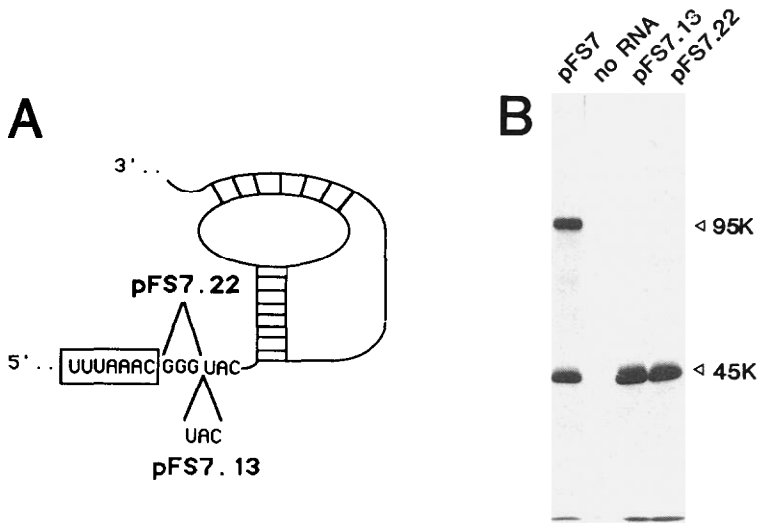

Figure 6. The Effect on Frameshifting Efficiency of Changing the Relative Position of the "Slippery" Sequence with Respect to the Tertiary RNA Structure

(A) Diagram showing two mutant derivatives of pFS7 in which the distance between the UUUAAAC sequence (boxed) and the tertiary RNA structure was increased (pFS7.13) or decreased (pFS7.22) by three nucleotides.

(B) Reticulocyte lysate translation products synthesized in response to mRNAs derived from SmaI-digested pFS7 or mutant templates. Polypeptides were labeled and analyzed as described in the legend to Figure 2.

at the downstream codon in the slippery sequence, i.e., at the leucine codon in RSV (AAAUUUA) and HIV-1 (UUUUUUA), and at the asparagine codon in MMTV *gag/pro* (AAAAAAC) (Hizi et al., 1987). It seems likely therefore that the IBV shift occurs at the asparagine codon (UUUAAAC), although confirmation will require protein sequencing at the junction of the frameshifted product.

Jacks et al. (1988b) have proposed a model for ribosomal frameshifting in RSV which involves simultaneous slippage on the mRNA of tRNAs occupying both the P and A sites of the ribosome. They postulate that such slippage would be possible only if base pairing could subsequently be maintained between the first two positions of each anticodon loop and the new mRNA codon. Our data for IBV frameshifting are consistent with such a model. Following slippage, base pairing could be maintained between the first two nucleotides of each anticodon loop and the mRNA (with the post-slippage anticodons AAU and UUG pairing with the mRNA codons UUU and AAA, respectively, after slippage). If this model is correct, it might be expected that reducing the capacity to form base pairs in the "slipped" position would affect frameshifting dramatically, and this indeed was observed. Two single site mutations (UUACAC, pFS7.14 and UUUAUAC, pFS7.17), which would prevent tRNA-mRNA pairing in the A site following slippage, greatly reduced or abolished frameshifting (0.5% or less). In mutant pFS7.1 (CUUAAAC), in which complementarity in the "slipped" position would be reduced from 2 to 1 bp in the P site, frameshifting was again markedly reduced, although in this case, some activity was retained (efficiency of 2%). This difference in frameshifting efficiency between the A and P site mutants may reflect the number of base pairs that can form following slippage at each site. Alternatively, the tRNA reading the A site may play a more important role in frameshifting than that occupying the P site, as suggested by Jacks and colleagues (1988b) for RSV and other retroviral frameshift sites.

Evidence for Tertiary RNA Folding in the IBV Frameshift Signal

Our deletion analysis of the frameshift signal of IBV indicates the requirement not only for the UUUAAAC "slippery sequence" at the 5′ end, but also of a stretch of nucleotides downstream. We previously suggested that a stem–loop structure predicted from the primary sequence of this downstream region might be an important element of the frameshift signal (Brierley et al., 1987), and the results reported here are consistent with this. However, our experiments indicate that this stem–loop forms part of a more complex tertiary structure, since base pairing was apparent between nucleotides of the predicted loop and a region downstream. Until direct structural information is obtained, the precise three-dimensional organization of the RNA remains uncertain. However, we believe that the RNA folds to form a structure similar in principle to the pseudoknots predicted to occur in a number of plant viral RNAs (see Pleij et al., 1985), foot and mouth disease virus (FMDV) RNA (Clarke et al., 1987), E. coli 5S RNA (Goringer and Wagner, 1986), E. coli 16S rRNA (Maly and Brimacombe, 1983; Pleij et al., 1985), and in the E. coli α mRNA leader sequence (Deckman and Draper, 1987). A similar kind of structure has also been described for a short synthetic RNA (Puglisi et al., 1988). The construction of the proposed IBV pseudoknot is outlined in Figure 5. A stretch of nucleotides in the loop of the hairpin and bordering the stem region form normal Watson–Crick base pairs with nucleotides of the complementary sequence located downstream in the RNA. Two separate double-helical regions form two stem structures (marked S1 and S2), which stack coaxially to generate an elongated double-stranded helix. It has been assumed (Pleij et al., 1985) that in interactions of this kind, the stems are stacked on top of each other such that a quasi-continuous, right-handed double helix is formed, comparable to A-RNA. Such an assumption is valid if the single stranded connecting loops L1 and L2 pose no sterical constraints upon

this structure (for a detailed discussion, see Pleij et al., 1985). The end product of the folding is a pseudoknot (Studnicka et al., 1978) in which the connecting loops L1 and L2 span the deep and shallow grooves of the double-helix, respectively.

The IBV stem–loop predicted by the computer-based RNA folding program of Jacobson et al. (1984) (Figure 3) has 11 bp in the stem and 5 unpaired nucleotides in the loop. As it stands, it seems unlikely that this precise structure could form a pseudoknot, since all the loop nucleotides are predicted to base pair with elements downstream, and no unpaired nucleotides remain to cross the deep groove of the double-stranded helix. To account for the results of our mutagenic analysis of the region, we suggest that an alternative stem–loop forms (shown in Figure 4) which, because a number of previously paired stem nucleotides now form part of the single-stranded loop, is capable of the tertiary interaction. In this model, the weak UG pair and the GC pair at the top of the stem are included in the loop (as is the previously unpaired G). This would potentially allow seven of the nine loop nucleotides to base pair with the downstream region to form S2, leaving two unpaired loop nucleotides to span the deep groove in the pseudoknot. The bridging of 7 bp in the deep groove by two loop nucleotides is sterically feasible (see Pleij, 1985). An artist's impression of the three-dimensional folding of the proposed IBV pseudoknot is shown in Figure 5. A quasi-continuous double helix of 16 bp in length is formed. As is the case in a number of the plant viral pseudoknots, loop 2 is considerably longer than loop 1. As only 7 bp can form between the loop and the downstream sequence, it is unlikely that a real knot could form in the RNA, since a full turn of an A-RNA duplex contains 11 bp (Arnott et al., 1972).

Our results therefore suggest strongly that an RNA pseudoknot is an essential element of the IBV frameshifting signal, although its precise organization remains to be elucidated. However through direct structural analysis of artificially synthesised wild-type and mutated mRNAs in combination with functional analysis by in vitro translation, it should be possible to investigate in considerable detail this novel RNA building principle.

The Role of Tertiary RNA Structure in Ribosomal Frameshifting

How does the presence of a pseudoknot in IBV mRNA lead to efficient frameshifting? The answer to this question is not yet clear, but an obvious possibility is that the elongating ribosome may be required to unwind the pseudoknot, and this could slow or stall its passage along the mRNA sufficiently to allow tRNA slippage at the "slippery" site. In this case it might be expected that mutations which destabilize, but do not abolish entirely, secondary and tertiary structure in the downstream region would be partially inhibitory to frameshifting. Indeed, in the case of mutants where the first predicted stem (S1) was allowed to form but not the second (S2), that is pFS7.5, pFS7.11, pFS7.12, and pFS7.15, frameshifting was detected, although at a much lower level. This raises the possibility, currently being tested, that intermediate levels of frameshifting would be observed if one or other of the two stems were only partially destabilized. Whatever model is advanced to explain how a pseudoknot promotes frameshifting must account for the precise spacing required between the knot and the "slippery" sequence, since deletion or insertion of as little as three bases was sufficient to reduce frameshifting dramatically. A possible explanation for this is that in these mutants, the ribosomal P and A sites are positioned on sequences that are refractory to slippage during a ribosomal pause.

A second important question arising out of this work is whether tertiary structures of this kind are a more general feature of (−1) frameshift sites. A large number of retroviruses, related retrotransposons (Jacks et al., 1988b) and a luteovirus (Miller et al., 1988) are known or suspected to utilize (−1) frameshifting. We have examined the known or suspected sites of such frameshifting (those listed and referenced in Jacks et al., 1988b) by predicting, with the aid of a computer program, the secondary structure of the RNA at the appropriate regions and then looking for potential tertiary base pairing interactions in the downstream sequences. We found that the "fold" program of Jacobson and colleagues (1984) predicted hairpin loops immediately downstream of the slippery sequences at all the sites except in the retrotransposon 17.6. We scanned the region downstream of the predicted stem–loops for sequences that were complementary to the loop nucleotides, setting 50 nucleotides as the maximum distance scanned and 5 nucleotides as the minimum size of the complementary region. The result obtained was rather striking: 14 out of the 22 sequences examined appear to contain the potential for pseudoknot formation. The number of complementary nucleotides was between 5 and 8, and these were located between 3 and 32 nucleotides from the base of the upstream stem–loop. Shown in Figure 7 are the potential interactions at the RSV *gag/pol*, MMTV *gag/pro* and *pro/pol* overlaps and some examples from cases where frameshifting is suspected to occur, with the complementary loop and downstream nucleotides highlighted. In those systems where frameshifting has been confirmed (RSV, MMTV, IBV, and HIV-1), only HIV-1 did not contain a suitable downstream sequence. Madhani et al. (1988) and Wilson et al. (1988), however, have observed recently that frameshifting in HIV-1 appears to occur in the absence of any recognizable RNA structures at the frameshift site, so this may not be unexpected. In addition to HIV-1, seven of the suspected sites did not show potential for the formation of tertiary RNA structure within the limits of this simple examination. Nevertheless, the possibility remains that tertiary RNA interactions may be a common feature of (−1) frameshift sites. In the case of RSV, it has been shown (Jacks et al., 1988b) that a deletion which left the stem–loop intact, but did not include downstream RSV sequences, had reduced frameshifting efficiency. However, when an additional 22 downstream nucleotides were included, frameshifting was restored to approximately wild-type efficiency. As can be seen in Figure 7, the additional sequences include an eight nucleotide stretch (UGUAGCGC) with full complementarity to a region in the loop of the RSV hairpin.

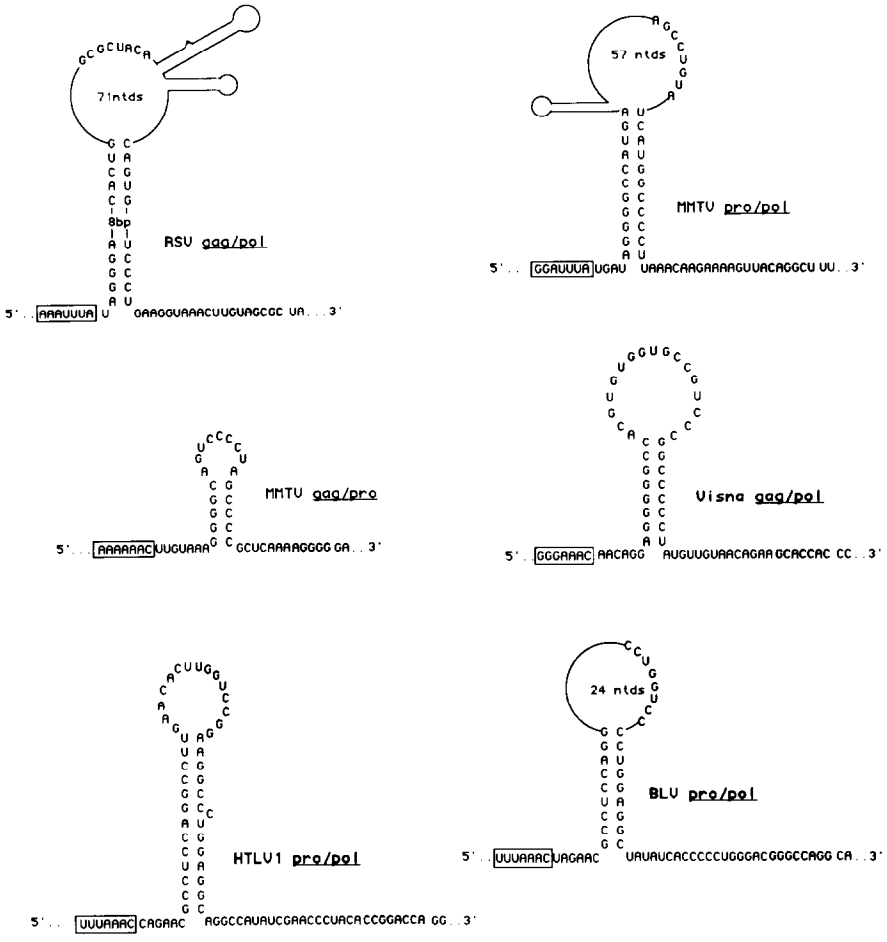

Figure 7. Potential Tertiary RNA Structures at Other (−1) Frameshift Sites

The figure shows the RNA secondary structures predicted by the "fold" program (Jacobson et al., 1984) at a number of retroviral gene overlaps where frameshifting is known (RSV, MMTV) or suspected to occur (Visna, HTLV1, BLV). Nucleotides (ntds) that could participate in base pairing interactions between the loop and downstream regions are highlighted. The known or suspected "slippery" sites are boxed. The nucleotide sequences were taken from the references cited in Jacks et al. (1988b).

which might therefore interact through base pairing. Introduction of single and compensatory base changes in this region, through site-directed mutagenesis, should resolve this question.

Experimental Procedures

Construction of Plasmid pFS7

Plasmid pFS1 (Brierley et al., 1987) was digested with DdeI and a 230 bp fragment containing the F1/F2 overlap region isolated (IBV sequence information from position 12,286 to 12,509 bp; Boursnell et al., 1987). Following end-filling with E. coli DNA polymerase I Klenow enzyme, the fragment was inserted into SspI partially digested plasmid pST1+ (Digard et al., submitted) to produce pFS6. pST1+ comprises the influenza virus A/PR8/34 PB1 gene (Young et al., 1983) inserted into the BglII site of plasmid pSP64-T (Krieg and Melton, 1984). Thus pFS6 consists of pSP64- T/PB1 with the IBV frameshift region inserted at position 1162 in the PB1 gene, such that both the F1 and F2 ORF were in frame with the flanking PB1 sequences, with F1 in frame 1 and F2 in frame 3. Plasmid pFS7 (see Figure 1) was constructed by inserting a 519 bp RsaI fragment from pEMBL8 (Dente et al., 1983) into the unique PvuII site of pFS6. This fragment included the intergenic region of filamentous bacteriophage f1 (Dotto et al., 1981) enabling single-stranded pFS7 DNA to be generated following infection of plasmid-carrying bacteria with helper bacteriophage R408 (Russel et al., 1986). The orientation of this fragment was such that the noncoding strand of pFS7 was packaged into R408 phage particles. The influenza-IBV-influenza junctions were confirmed by dideoxy nucleotide sequencing (Sanger et al., 1977) of single-stranded pFS7. Both pFS6 and pFS7 were designed such that ribosomes could translate the length of the PB1 mRNA following a frameshift event within the inserted IBV region (see Figure 1). Plasmids were maintained in E. coli JM101 (Yanisch-Perron et al., 1985).

Site-Specific Mutagenesis of pFS7

A procedure based on the method of Kunkel (1985) was employed to prepare site-specific mutations within the IBV frameshift region. Plasmid pFS7 (or mutant derivatives) was transformed into E. coli strain RZ1032 (dut^-ung^-) (Kunkel, 1985) and single-stranded, uracil-containing, pFS7 DNA isolated following superinfection with bacteriophage R408. Phosphorylated mutagenic oligonucleotides (5 pmol) were annealed to this template (0.1–0.5 μg) in a 10 μl reaction mixture containing 10 mM Tris-HCl (pH 8), 10 mM $MgCl_2$, 50 mM NaCl, and 1 mM dithiothreitol (DTT) at 55°C for 5 min. After cooling at room temperature (10 min), the primer was extended by addition of dATP, dCTP, dGTP, and dTTP (0.5 mM each); ATP (0.5 mM), E. coli DNA polymerase I Klenow enzyme (6 U, Boehringer Mannheim) and T4 DNA ligase (2.5 U, Bethesda Research Laboratories) to a final reaction volume of 20 μl after readjusting the concentration of Tris-HCl, $MgCl_2$, and DTT to 40 mM, 7.5 mM, and 3 mM, respectively. The extension–ligation reac-

tion was for 2 hr at room temperature, and then for an additional 2 hr after supplementing the reaction with additional Klenow (6 U) and ligase (2.5 U). Portions of the reaction were transformed into E. coli JM101, and single-stranded templates prepared from transformants by R408 superinfection. Mutants were identified by dideoxy sequencing. The efficiency of mutagenesis was between 20% and 70%.

Preparation of DNA Template, Transcription and Translation of mRNA

Plasmids for transcription were prepared by the alkaline lysis minipreparation method (Birnboim and Doly, 1979), linearized by digestion with SmaI, and transcribed as described previously (Brierley et al., 1987). Messenger RNAs were translated in rabbit reticulocyte lysates as described before (Brierley et al., 1987). Transcripts were translated in Xenopus oocytes as described by Colman (1984). Individual oocytes were microinjected with approximately 50 ng mRNA, and labeled with [^{35}S]methionine (10 μCi) from 2–6 hr following injection. Translation products were analyzed on 10% SDS–polyacrylamide gels according to standard procedures (Hames, 1981). The relative abundance of nonframeshifted or frameshifted products on the gels was estimated by either scanning densitometry or by measuring the radioactive content of excised gel slices. In each case, a correction was made to account for the differential methionine content of the products.

Antisera against Defined Regions of the Influenza Virus PB1 Protein

Rabbit antisera against defined regions of the PB1 protein were prepared by immunization with PB1 fragments expressed as fusion proteins with β-galactosidase (Stanley and Luzio, 1984). Antisera 1 (as1) and 4 (as4) covered the regions encoded by nucleotides 150–1110 and 1464–2256, respectively. An antiserum raised against a region of the influenza A/PR8/34 PA protein by a similar procedure was used as a control. Radioimmune precipitation of reticulocyte translation products was performed as described (Brierley et al., 1987).

Acknowledgments

We are grateful to Bhavna Bilimoria and Nicola Rolley for expert assistance, to Mike Boursnell for useful discussion, and to Tom Atkinson and Professor Michael Smith for the gift of oligonucleotides. This work was supported by a grant from the Agricultural Research Council (LRG 45).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received February 2, 1989; revised April 7, 1989.

References

Arnott, S., Hukins, D. W. L., and Dover, S. D. (1972). Optimised parameters for RNA double-helices. Biochem. Biophys. Res. Commun. *48*, 1392–1399.

Birnboim, H. C., and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl. Acids Res. *7*, 1513–1523.

Boursnell, M. E. G., Brown, T. D. K., Foulds, I. J., Green, P. F., Tomley, F. M., and Binns, M. M. (1987). Completion of the sequence of the genome of the coronavirus avian infectious bronchitis virus. J. Gen. Virol. *68*, 57–77.

Brierley, I., Boursnell, M. E. G., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., and Inglis, S. C. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. *6*, 3779–3785.

Clare, J. J., and Farabaugh, P. J. (1985). Nucleotide sequence of a yeast Ty element: evidence for an unusual mechanism of gene expression. Proc. Natl. Acad. Sci. USA *82*, 2829–2833.

Clare, J. J., Belcourt, M., and Farabaugh, P. J. (1988). Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. Proc. Natl. Acad. Sci. USA *85*, 6816–6820.

Clarke, B. E., Brown, A. L., Curry, K. M., Newton, S. E., Rowlands, D. J., and Carroll, A. R. (1987). Potential secondary and tertiary structure in the genomic RNA of foot and mouth disease virus. Nucl. Acids Res. *15*, 7067–7080.

Colman, A. (1984). Translation of eukaryotic messenger RNA in Xenopus oocytes. In Transcription and Translation: A Practical Approach, B. D. Hames and S. J. Higgins, eds. (Oxford: IRL Press), pp. 271–302.

Craigen, W. J., and Caskey, C. T. (1987). Translational frameshifting: where will it stop? Cell *50*, 1–2.

Deckman, I. C., and Draper, D. E. (1987). S4-α mRNA translation regulation complex. II. Secondary structures of the RNA regulatory site in the presence and absence of S4. J. Mol. Biol. *196*, 323–332.

Dente, L., Cesareni, G., and Cortese, R. (1983). pEMBL: a new family of single-stranded plasmids. Nucl. Acids Res. *11*, 1645–1655.

Dotto, G. P., Enea, V., and Zinder, N. D. (1981). Functional analysis of bacteriophage f1 intergenic region. Virology *114*, 463–473.

Goringer, H. U., and Wagner, R. (1986). Does 5S RNA from E. coli have a pseudoknotted structure? Nucl. Acids Res. *14*, 7473–7485.

Hames, B. D. (1981). An introduction to polyacrylamide gel electrophoresis. In Gel Electrophoresis of Proteins—A Practical Approach, B. D. Hames and D. Rickwood, eds. (Oxford: IRL Press), pp. 1–91.

Hizi, A., Henderson, L. E., Copeland, T. D., Sowden, R. C., Hixson, C. V., and Oroszlan, S. (1987). Characterization of mouse mammary tumor virus *gag-pol* gene products and the ribosomal frameshift by protein sequencing. Proc. Natl. Acad. Sci. USA *84*, 7041–7046.

Hodgman, T. C. (1988). A new superfamily of replicative proteins. Nature *333*, 22–23.

Jacks, T., and Varmus, H. E. (1985). Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science *230*, 1237–1242.

Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987). Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins. Proc. Natl. Acad. Sci. USA *84*, 4298–4302.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988a). Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. Nature *331*, 280–283.

Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988b). Signals for ribosomal frameshifting in the Rous sarcoma virus *gag-pol* region. Cell *55*, 447–458.

Jacobson, A. B., Good, L., Simonetti, J., and Zuker, M. (1984). Some simple computational methods to improve the folding of large RNAs. Nucl. Acids Res. *12*, 54–62.

Krieg, P. A., and Melton, D. A. (1984). Functional messenger RNAs are produced by SP6 in vitro transcription of cloned cDNAs. Nucl. Acids Res. *12*, 7057–7071.

Kunkel, T. A. (1985). Rapid and efficient site-specific mutagenesis without phenotypic selection. Proc. Natl. Acad. Sci. USA *82*, 488–492.

Madhani, H. D., Jacks, T., and Varmus, H. E. (1988). Signals for the expression of the HIV *pol* gene by ribosomal frameshifting. In The Control of HIV Gene Expression, R. Franza, B. Cullen, and F. Wong-Staal, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 119–125.

Maly, P., and Brimacombe, R. (1983). Refined secondary structure models for the 16S and 23S ribosomal RNA of Escherichia coli. Nucl. Acids Res. *11*, 7263–7286.

Marlor, R. L., Parkhurst, S. M., and Corces, V. G. (1986). The Drosophila melanogaster gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell Biol. *6*, 1129–1134.

Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J. (1985). A retrovirus-like strategy for expression of a fusion protein encoded by yeast transposon Ty1. Nature *313*, 243–246.

Miller, W. A., Waterhouse, P. M., and Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. Nucl. Acids Res. *16*, 6097–6111.

Moore, R., Dixon, M., Smith, R., Peters, G., and Dickson, C. (1987). Complete nucleotide sequence of a milk-transmitted mouse mammary

tumor virus: two frameshift suppression events required for translation of *gag* and *pol*. J. Virol. *61*, 480–490.

Pleij, C. W. A., Rietveld, K., and Bosch, L. (1985). A new principle of RNA folding based on pseudoknotting. Nucl. Acids Res. *13*, 1717–1731.

Puglisi, J. D., Wyatt, J. R., and Tinoco, I. (1988). A pseudoknotted RNA oligonucleotide. Nature *331*, 283–286.

Rice, N. R., Stephens, R. M., Burny, A., and Gilden, R. V. (1985). The *gag* and *pol* genes of bovine leukemia virus: nucleotide sequence and analysis. Virology *142*, 357–377.

Rietveld, K., Pleij, C. W. A., and Bosch, L. (1983). Three-dimensional models of the tRNA–like 3′-termini of some plant viral RNAs. EMBO J. *2*, 1079–1085.

Roth, J. R. (1981). Frameshift suppression. Cell *24*, 601–602.

Russel, M., Kidd, S., and Kelley, M. R. (1986). An improved filamentous helper phage for generating single-stranded plasmid DNA. Gene *45*, 333–338.

Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., and Yuki, S. (1984). Identification of the coding sequence for a reverse transcriptase–like enzyme in a transposable genetic element in Drosophila melanogaster. Nature *312*, 659–661.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA *74*, 5463–5467.

Schochetman, G., Stevens, R. H., and Simpson, R. W. (1977). Presence of infectious polyadenylated RNA in the coronavirus avian infectious bronchitis virus. Virology *77*, 772–782.

Siddell, S., Wege, H., and ter Meulen, V. (1983). The biology of coronaviruses. J. Gen. Virol. *64*, 761–776.

Stanley, K. K., and Luzio, J. P. (1984). Construction of a new family of high efficiency bacterial expression vectors: identification of cDNA clones coding for human liver proteins. EMBO J. *3*, 1429–1434.

Stern, D. F., and Kennedy, S. I. T. (1980). Coronavirus multiplication strategy. I. Identification and characterization of virus-specified RNA. J. Virol. *34*, 665–674.

Studnicka, G. M., Rahn, G. M., Cummings, I. W., and Salser, W. A. (1978). Computer method for predicting the secondary structure of single-stranded RNA. Nucl. Acids Res. *5*, 3365–3387.

Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J. and Kingsman, S. M. (1986). Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. Nucl. Acids Res. *14*, 7001–7015.

Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J. (1988). HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell *55*, 1159–1169.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13 mp18 and pUC19 vectors. Gene *33*, 103–119.

Young, J. F., Desselberger, U., Graves, P., Palese, P., Shatzman, A., and Rosenberg, M. (1983). Cloning and expression of influenza virus genes. In The Origin of Pandemic Influenza Viruses, W. G. Laver, ed. (Amsterdam: Elsevier Science), pp. 129–138.