



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

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

Signals for Ribosomal Frameshifting in the Rous Sarcoma Virus *gag-pol* Region

Tyler Jacks,* Hiten D. Madhani,* Frank R. Masiarz,† and Harold E. Varmus*

* Department of Biochemistry and Biophysics and Department of Microbiology and Immunology University of California

San Francisco, California 94143

† Chiron Corporation

Emeryville, California 94608

Summary

The *gag-pol* protein of Rous sarcoma virus (RSV), the precursor to the enzymes responsible for reverse transcription and integration, is expressed from two genes that lie in different translational reading frames by ribosomal frameshifting. Here, we localize the site of frameshifting and show that the frameshifting reaction is mediated by slippage of two adjacent tRNAs by a single nucleotide in the 5' direction. The *gag* terminator, which immediately follows the frameshift site, is not required for frameshifting. Other suspected retroviral frameshift sites mediate frameshifting when placed at the end of RSV *gag*. Mutations in RSV *pol* also affect synthesis of the *gag-pol* protein in vitro. The effects of these mutations best correlate with the potential to form an RNA stem-loop structure adjacent to the frameshift site. A short sequence of RSV RNA, 147 nucleotides in length, containing the frameshift site and stem-loop structure, is sufficient to direct frameshifting in a novel genetic context.

Introduction

The vast majority of eukaryotic mRNAs are monocistronic (Kozak, 1987). Unlike their prokaryotic counterparts, eukaryotic ribosomes tend not to initiate at internal methionine codons, and thus translation on a eukaryotic mRNA is usually limited to the open reading frame that follows the first AUG codon (Kozak, 1978). Consequently, there are very few examples of coordinate synthesis of multiple protein products from individual eukaryotic mRNA species (Kozak, 1986). There is an emerging class of eukaryotic mRNAs, however, that do encode multiple proteins, not by controlling where ribosomes begin translating but where they finish, either by the suppression of in-frame termination codons or by ribosomal frameshifting (Clare and Farabaugh, 1985; Mellor et al., 1985; Yoshinaka et al., 1985a, 1985b; Jacks and Varmus, 1985; Moore et al., 1987; Jacks et al., 1987, 1988; Brierley et al., 1987).

In all known retroviruses, the *pol* gene (encoding the reverse transcriptase and integrase functions) lies downstream of the *gag* gene, which codes for the virus core proteins (Weiss et al., 1982). Retroviruses arrange their *gag* and *pol* genes in one of three ways: in the same reading frame, separated by a single termination codon; in different reading frames, with *pol* briefly overlapping *gag* in the -1 direction; or with a third gene (encoding the viral pro-

tease, termed *pro*) between *gag* and *pol* and overlapping both (Varmus, 1988). Despite these apparent blocks to continuous translation, all retroviruses initially express *pol* by first synthesizing a *gag-pol* (or *gag-pro-pol*) fusion protein that is later cleaved during virus assembly to yield the mature products. The ratio of this fusion protein to the product of the *gag* gene alone is approximately 1:20 (Weiss et al., 1982).

Yoshinaka et al. (1985a) first showed by direct amino acid sequencing that the termination codon separating the murine leukemia virus (MLV) *gag* and *pol* genes is efficiently suppressed by a glutamine-charged tRNA. In vitro transcription and translation methods were then used to demonstrate ribosomal frameshifting during expression of the *gag-pol* protein of Rous sarcoma virus (RSV) (Jacks and Varmus, 1985) and human immunodeficiency virus type 1 (HIV-1) (Jacks et al., 1988) and double frameshifting in the synthesis of the mouse mammary tumor virus (MMTV) *gag-pro-pol* protein (Jacks et al., 1987; Moore et al., 1987). The genomic sequences of several other retroviruses indicate that they utilize one of these three strategies to express their *pol* genes (Varmus, 1988).

In this report, we examine the sequence requirements for ribosomal frameshifting during translation of retroviral RNAs, using RSV as a model system. Radiolabeled amino acid sequencing and site-directed mutagenesis were used to localize the precise site of frameshifting in RSV RNA to the last *gag* codon, a UUA-leucine, and suggest that the -1 frameshift is mediated by the simultaneous slippage of two tRNAs, the UUA-reading tRNA^{Leu} and the one preceding it, by one nucleotide in the 5' direction. Certain other sequences will functionally substitute for the natural RSV sequence at the frameshift site, including the sequences A AAA AAC and U UUA AAC, which are suspected to be the sites of frameshifting in other retroviral RNAs.

However, these signals appear insufficient since we have previously shown that certain retroviral frameshift sites fail to allow ribosomal frameshifting in a heterologous genetic context. We have proposed that potential stem-loop structures positioned just downstream of all known or suspected frameshift sites are a second necessary element in the frameshifting process (Jacks et al., 1987, 1988). We now demonstrate by deletion and site-directed mutagenesis of the 5' region of RSV *pol* gene that a stem-loop structure is required for efficient frameshifting in vitro for this virus. In addition, a 147 nucleotide sequence containing the RSV frameshift site and stem-loop structure is sufficient to cause efficient ribosomal frameshifting when placed in a heterologous context.

Results

Common Sequences within Different Retroviral Overlap Regions

Site-specific frameshifting within the various retroviral overlap regions was first suggested by the observation that each of these regions contains one of three common

Table 1. Demonstrated and Suspected Retroviral Frameshift Sites

Virus or Transposable Element	Overlap	Sequence	Distance from 3' End of Overlap (nucleotides)
RSV	<i>gag/pol</i>	ACA AAU UUA UAG	0
HIV-1	<i>gag/pol</i>	AAU UUU UUA GGG	198
HIV-2	<i>gag/pol</i>	GGU UUU UUA GGA	267
SIV	<i>gag/pol</i>	GGU UUU UUA GGC	213
<i>gypsy</i>	<i>gag/pol</i>	AAU UUU UUA GGG	51
MMTV	<i>pro/pol</i>	CAG GAU UUA UGA	0
SRV-1	<i>pro/pol</i>	GGA AAU UUU UAA	0
MPMV	<i>pro/pol</i>	GGA AAU UUU UAA	0
17.6	<i>gag/pol</i>	GAA AAU UUU CAG	30
Mouse IAP	<i>gag/pol</i>	CUG GGU UUU CCU	3
MMTV	<i>gag/pro</i>	UCA AAA AAC UUG	3
BLV	<i>gag/pro</i>	UCA AAA AAC UAA	0
HTLV-1	<i>gag/pro</i>	CCA AAA AAC UCC	18
HTLV-2	<i>gag/pro</i>	GGA AAA AAC UCC	18
EIAV	<i>gag/pol</i>	CCA AAA AAC GGG	195
BLV	<i>pro/pol</i>	CCU UUA AAC UAG	0
HTLV-1	<i>pro/pol</i>	CCU UUA AAC CAG	156
HTLV-2	<i>pro/pol</i>	CCU UUA AAC CUG	18
SRV-1	<i>gag/pro</i>	CAG GGA AAC GAC	147
MPMV	<i>gag/pro</i>	CAG GGA AAC GGG	147
Visna	<i>gag/pol</i>	CAG GGA AAC AAC	45

Common heptanucleotide sequence motifs are present in all retroviral overlaps known or presumed to contain sites of frameshifting. The heptanucleotides are shown in boldface type along with their neighboring sequences and the distance (in nucleotides) between the 3' nucleotide of the heptameric sequence and the 3' end of its overlap (as delineated by the first nucleotide of the 0-frame termination codon). Sequences are grouped according to their final three nucleotides; these constitute a codon in the upstream (e.g., *gag*) gene. Two of these codons, UUA and AAC, have previously been identified as the sites of frameshifting (see text). Evidence that the entire heptanucleotide sequence may participate in the frameshifting reaction is presented in this report. References for nucleotide sequences: RSV (Schwartz et al., 1983); HIV-1 (Wain-Hobson et al., 1985; Ratner et al., 1985; Sanchez-Pescador et al., 1985); HIV-2 (Guyader et al., 1987); simian immunodeficiency virus (SIV) (Chakrabarti et al., 1987); *gypsy* (Marlor et al., 1986); MMTV (Jacks et al., 1987; Moore et al., 1987); SRV-1 (Power et al., 1986); Mason-Pfizer monkey virus (MPMV) (Sonigo et al., 1986); 17.6 (Saigo et al., 1984); mouse intracisternal A particle (IAP) (Meitz et al., 1987); bovine leukemia virus (BLV) (Sagata et al., 1985; Rice et al., 1985); human T cell leukemia virus type 1 (HTLV-1) (Hiramatsu et al., 1987) and type 2 (HTLV-2) (Shimotohno et al., 1985); equine infectious anemia virus (EIAV) (Stephens et al., 1986); and visna virus (Sonigo et al., 1985).

sequences. (The overlap regions are delineated on the 3' side by the termination codon of the upstream [e.g., *gag*] open reading frame and on the 5' side by the termination codon that demarcates the beginning of the downstream open reading frame [e.g., *pol*]). As shown in Table 1, several overlaps, including the *gag/pol* overlaps of RSV and HIV-1, contain the sequence U UUA (where the UUA is a leucine codon in the 0 frame). One of two other sequences, U UUU or A AAC, appears in each of the remaining overlaps (Table 1). We have previously proposed that -1 frameshifting might occur at these sequences if the tRNAs reading the 0-frame codons occasionally slipped back and paired with the codons in the -1 frame (Jacks et al., 1988). In fact, amino acid sequencing has shown that two of these sequences, U UUA and A AAC, are the sites of frameshifting during HIV-1 *gag-pol* (Jacks et al., 1988) and MMTV *gag-pro* expression (Hizi et al., 1987), respectively, and the amino acid sequences are consistent with the proposed mechanism.

The sequence similarity among the different retroviral overlaps actually extends 5' to the putative frameshift sites. Table 1 shows that in all but one of the overlaps, the three common sequences are preceded by runs of three U, A, or G residues, creating similar sequence motifs that are seven nucleotides in length. (The U UUA sequence in

the MMTV *pro-pol* overlap, the sole exception, is preceded by the sequence GGA.) Similar heptameric sequences are found in the *gag/pol* overlaps of two retrotransposons of *Drosophila*, 17.6 and *gypsy*, and the mouse intracisternal A particle (Table 1).

The Simultaneous-Slippage Model of Frameshifting

One unifying explanation for the observed arrangement of nucleotides at and upstream of the putative frameshift sites is that two adjacent ribosome-bound tRNAs are required to slip into the -1 frame during retroviral frameshifting. This model is shown in some detail for the RSV sequence A AAU UUA in Figure 1. In step I, normal translation delivers a ribosome to the final two codons of *gag* such that the UUA codon is in the ribosomal A site being read by tRNA^{Leu}. The nascent protein is carried by the tRNA^{Asn} reading the AAU codon in the P site. Simultaneous slippage of these two tRNAs by one nucleotide in the 5' direction leads to the conformation shown in step II, where both tRNAs are base paired to the mRNA in two out of three anticodon positions. This interaction is made possible by the A and U residues 5' to the AAU and UUA codons, respectively. Next, normal peptidyl transfer of the nascent protein to the tRNA^{Leu} and translocation of this tRNA to the P site brings the first *pol*-frame codon (AUA)

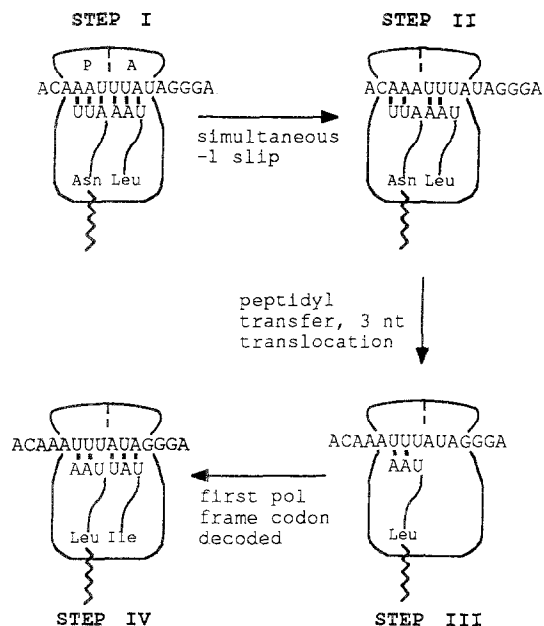


Figure 1. The Simultaneous-Slippage Model of Frameshifting at the RSV Frameshift Site

tRNA^{Asn} carrying the nascent peptide (jagged line) and tRNA^{Leu} are shown bound to the *gag*-frame codons, AAU and UUA, in the ribosomal P and A sites (step I). Simultaneous slippage of the two tRNAs by one nucleotide in the 5' direction results in their complexing with the adjacent *pol*-frame codons, AAA and UUU, with base pairs (bars) in the first and second codon positions (step II). Normal peptidyl transfer and three-nucleotide translocation brings the next *pol*-frame codon, AUA, into the A site (step III), where it is decoded by tRNA^{Ile} (step IV). Note that the slippage could also occur following peptidyl transfer and prior to translocation. Furthermore, sequences of the tRNA anticodons shown in this model are based on standard Watson-Crick base pairs. The actual anticodon sequences are not known (see text).

into the A site (step III), where it is normally decoded by tRNA^{Ile} (step IV). The other suspected frameshift sites listed in Table 1 would allow slippage by these or other tRNA species in a similar manner.

Amino Acid Sequencing at the *gag-pol* Junction

As a first step in testing the model presented in Figure 1, we used amino acid sequencing to demonstrate that the proposed frameshift site is, in fact, the point where ribosomes begin translation in the *pol* frame. We replaced nearly the entire RSV *gag* gene with an initiator methionine and two additional codons such that the position of translational initiation is just ten codons from the proposed site of frameshifting (Figure 2A). This plasmid, pGP-S, also has a portion of the *Staphylococcus aureus* protein A gene replacing the carboxyl terminus of RSV *pol* in order to facilitate purification of the resulting "trans-frame" protein (the product of frameshifting) using IgG-Sepharose. pGP-S retains the *pol* nucleotides predicted to form the stem-loop discussed below.

If the *gag*-frame UUA-leucine codon is the site of frameshifting, translation of pGP-S-encoded mRNA should proceed normally until the ribosomes decode this codon

(the eleventh). The simultaneous-slippage model (Figure 1) calls for the *pol*-frame AUA-isoleucine codon to be the next decoded. Thus, the amino acid sequence of IgG-Sepharose-purified material from translation of GP-S RNA should include leucine at position 11 and isoleucine at position 12. If frameshifting on GP-S RNA occurs upstream of the UUA codon, the eleventh decoded triplet would be the *pol*-frame UUU-phenylalanine (Figure 2A), and the resulting transframe protein would contain phenylalanine at position 11 (Figure 2A). Because UUA-leucine is the last codon in the overlap (Figure 2A), productive frameshifting cannot occur downstream of this site.

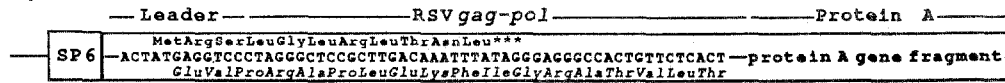
The histograms shown in Figure 2C display the amounts of radioactivity present in the first 20 cycles of Edman degradation of purified, pGP-S-encoded transframe protein synthesized in vitro in the presence of [³⁵S]methionine and either [³H]leucine (panel I), [³H]isoleucine (panel II), or [³H]phenylalanine (panel III). The peaks of radioactive leucine and isoleucine at positions 11 and 12, respectively, and the lack of radioactive phenylalanine confirm that the site of frameshifting is the terminal *gag* codon, UUA. The other observed peaks correspond to the methionine residue at position one and leucine residues at positions 4, 6, and 8 in the *gag* frame and position 18 in the *pol* frame (Figure 2A and B). Thus, the amino acid sequence encoded at the frameshift site is consistent with the proposed model (Figure 1).

Site-Directed Mutations in the Frameshift Site

While the determined amino acid sequence supports the simultaneous-slippage model, other models (for example, two-nucleotide translocation by the tRNA^{Leu} or slippage of this tRNA while resident in the ribosomal P site) would predict the same amino acid sequence. To test our model more directly, we constructed a series of site-directed mutations in and around the RSV frameshift site. To facilitate discussion of these mutations, the nucleotide positions have been numbered as shown in Figure 3A. (The first position of the UUA codon is designated +1, with positive and negative integers proceeding 3' and 5', respectively.) The mutations were constructed by oligonucleotide-directed mutagenesis (see Experimental Procedures) in an SP6 promoter-containing plasmid carrying the complete RSV *gag* gene and about one-third of the *pol* gene; frameshifting was assayed by the ability of RNAs transcribed from these mutants to direct synthesis of a 108 kd *gag-pol* fusion protein in a rabbit reticulocyte lysate in vitro translation system.

According to the simultaneous -1 slippage model (Figure 1), the seven nucleotides extending from the A residue at position -4 through to the A residue at position +3 (Figure 3A) participate in the frameshift event as part of the 0 and -1 frame codons bound by the frameshift-mediating tRNAs. The model predicts that mutations in these seven positions may be inhibitory. Conversely, the nucleotides neighboring the heptameric sequence play no obvious role in this mechanism; thus, mutations in these positions should be silent.

A.



pGP-S

B.

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20

HN-MetArgSerLeuGlyLeuArgLeuThrAsnLeu

2 *GluValProArgAlaProLeuGluLysPheIleGlyArgAlaThrValLeuThrVal* Protein A ccc⁻

C.

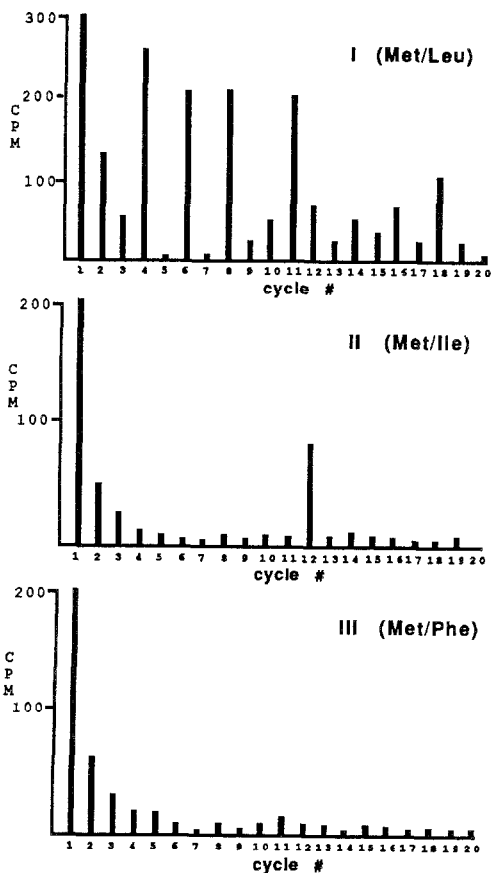


Figure 2. Amino Acid Sequence at the RSV Frameshift Site

(A) A portion of the protein sequencing vector pGP-S. Downstream of the SP6 promoter (Melton et al., 1984) was cloned a sequence composed of an initiator methionine codon and two additional codons (Arg and Ser) followed in frame by the 3' end of the RSV gag gene, beginning with the leucine codon located seven codons upstream of the gag terminator. (The amino terminus of the protein encoded by RNA transcribed from pGP-S, Met-Arg-Ser-Leu, is not acetylated in the rabbit reticulocyte lysate system [Jacks et al., 1988].) Following approximately 250 nucleotides of RSV pol sequence in pGP-S is a segment of the S. aureus protein A gene (Uhlen et al., 1983) in frame with pol. Thus, the transframe protein encoded by GP-S RNA is readily purified using IgG-Sepharose (Nilsson et al., 1985). The nucleotide sequence of the 5' end of this hybrid gene is shown along with the translation in the gag frame (above the nucleotide sequence) and pol frame (below the sequence and in italics).

(B) Possible amino-terminal amino acid sequences of the transframe protein synthesized from GP-S RNA. The amino acid sequences encoded by both the gag and pol frames are shown, with the pol sequence below and in italics. Amino acid positions are numbered.

(C) Histograms recording the amount of radioactivity present in the first 20 cycles of Edman degradation performed on IgG-Sepharose-purified protein synthesized from GP-S RNA in a rabbit reticulocyte lysate supplemented with [³⁵S]methionine and either [³H]leucine (panel I), [³H]isoleucine (panel II), or [³H]phenylalanine (panel III). CPM refers to counts per minute above background.

Mutations in Positions -1 through +2 Abolish Frameshifting: Support for tRNA^{Leu} Slippage

Slippage by the tRNA^{Leu} from the gag frame into the pol frame requires the integrity of the run of three U residues in positions -1 to +2. A mutation in the 5'-most U residue would impair the ability of the tRNA^{Leu} to slip back. Mutation of the following two U residues would change the 0-frame codon and thereby specify a tRNA that would be less likely to slip back given a U at position -1. As shown in Figure 3B, mutation of any of these three U residues to any other nucleotide severely inhibits frameshifting efficiency. Translation of RNAs carrying these mutations results in undetectable amounts of the gag-pol protein

(Figure 3B). Thus, -1 slippage of the tRNA^{Leu} is indicated.

Frameshifting is Not Affected by Mutations in the gag Terminator

If the tRNA^{Leu} slips into the pol frame while resident in the P site (Figure 1), mutations in the upstream AAU-asparagine codon (and the A preceding it) will adversely affect frameshifting and those in the downstream gag termination codon will not.

The proximity of the gag terminator to the frameshift site is a provocative finding, especially in light of the enhancement of frameshifting in Escherichia coli by 3' neighboring stop codons (Weiss et al., 1987). However, whether the

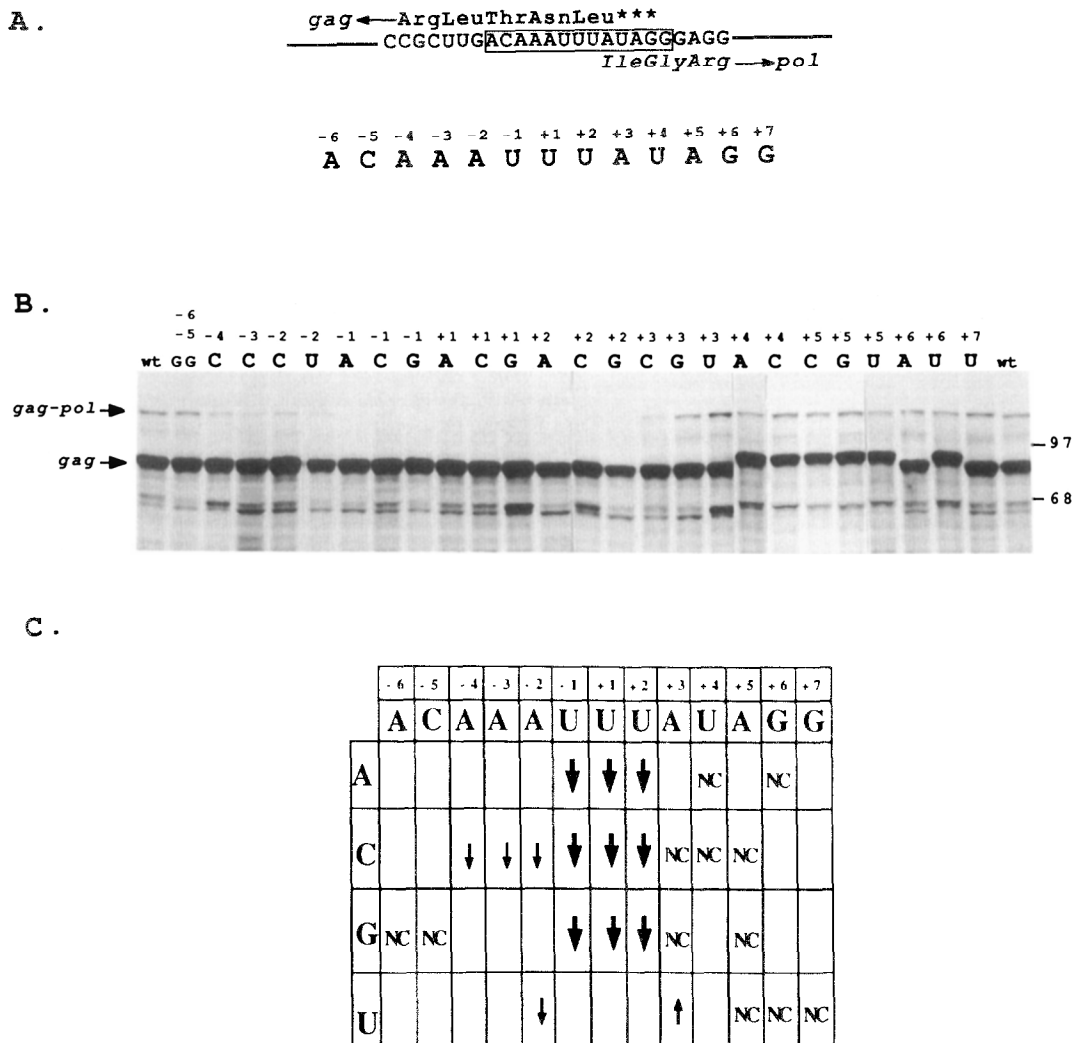


Figure 3. Effects of Point Mutations in and around the RSV Frameshift Site

(A) Nucleotides in the boxed region at the end of the RSV *gag* gene are designated numerically below the representation of the frameshift site. (B) Fluorogram of a 10% SDS-polyacrylamide gel containing total ³⁵S-labeled products of rabbit reticulocyte lysate translations directed by either wild-type (wt) or various mutant RSV RNAs. The specific mutations are indicated by their position (according to part A, above) and nucleotide change. The positions of the *gag* and *gag-pol* proteins and molecular mass markers (in kd, at right) are indicated. (C) Summary of mutational effects. The wild-type RSV sequence is shown horizontally, and possible base changes are listed vertically. Narrow and thick downward arrows indicate decreases in frameshifting efficiency of ~5-fold and >10-fold, respectively. A narrow, upward arrow indicates a 2-fold increase in frameshifting efficiency. NC symbolizes no change in efficiency. Frameshifting efficiencies were calculated from the amount of radioactivity in excised gel slices containing the *gag* and *gag-pol* proteins after correction for differential methionine content. The estimates reflect data collected from at least three separate experiments. Blank entries indicate either that mutations were not constructed or that the nucleotide corresponds to the wild-type sequence.

gag terminator was changed to a sense codon (positions +4 and +5 mutations and +6U), another stop codon (+6A), or was followed by a second stop codon (+7U), the observed frameshifting efficiency was unchanged (Figures 3B and 3C). Note that those mutations that convert the *gag* terminator to a sense codon extend the *gag* open reading frame by 111 nucleotides (Schwartz et al., 1983), resulting in a larger *gag* protein (Figure 3B). For at least one of these mutants (+6U), however, the site of frameshifting is unchanged, as determined from amino acid sequence analysis using a pGP-S derivative (see Figure 2A.) carrying this mutation (not shown).

Mutations in Positions -4 through +2 Inhibit Frameshifting, and a Mutation Further Upstream Does Not

The proposed model (Figure 1) stipulates that frameshifting on RSV RNA occurs while the tRNA^{Leu} is in the ribosomal A site and involves the simultaneous slippage of this tRNA and the P-site tRNA^{Asn}. These aspects of the model are most strongly supported by the reduction in frameshifting efficiency observed upon mutation of the three A residues in positions -4 to -2. Converting any of these A residues to C (-4C, -3C, and -2C) reduces frameshifting efficiency from the wild-type value of 5% to

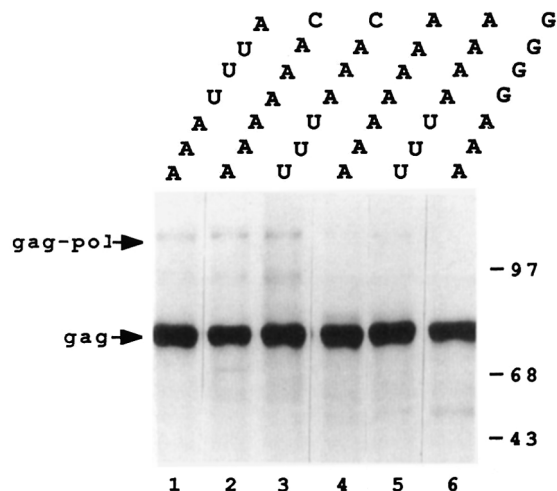


Figure 4. Frameshifting on Functional and Defective Replacements of the RSV Frameshift Site

Fluorogram of a 10% SDS-polyacrylamide gel containing total ³⁵S-labeled products of a rabbit reticulocyte lysate translation of wild-type RSV RNA (lane 1) or various mutant derivatives (lanes 2-6). The mutations affect all or a part of the RSV frameshift site (the last seven nucleotides of the *gag* gene). Sequences of the RSV frameshift site and mutant sites are shown above the lanes. The positions of the *gag* and *gag-pol* proteins and molecular mass standards (in kd, at right) are indicated.

approximately 1% (Figures 3B and 3C). The -2U mutation has a similar inhibitory effect. As with the inhibition caused by mutations in the run of U residues, we attribute these deleterious effects to the specification of a tRNA with a decreased probability of slipping back (position -3 and -2 mutations) or to the disruption of the site to which the tRNA^{Asn} normally slips (-4C).

The nucleotides 5' to the -4 position should not influence frameshifting according to the simultaneous-slippage model (Figure 1). Indeed, conversion of the wild-type AC dinucleotide at positions -6 and -5 to GG does not alter the ratio of the *gag* to *gag-pol* protein (-6G -5G, Figures 3B and 3C).

Mutations Are Tolerated in the +3 Position: Multiple tRNAs Can Mediate Frameshifting

In addition to favorable -1-frame base pairing by the tRNA^{Leu} and tRNA^{Asn}, we considered the possibility that frameshifting might require the action of specialized isoacceptor tRNA species having the unusual ability to slip into the -1 frame. The effects of the three mutations in the +3 position suggest that if such a requirement exists, it is not absolute. The mutations that convert the UUA-leucine codon to either UUG-leucine or UUC-phenylalanine still allow efficient frameshifting. (The frameshift efficiencies of the +3G and +3C mutants are 5% and 3%; Figures 3B and C). The +3U mutation (creating a UUU phenylalanine codon) actually enhances the frameshifting efficiency to 10%, twice the wild-type value.

More efficient frameshifting with the +3U mutant could mean that base-pairing potential in the -1 frame is solely responsible for how often tRNAs and, consequently, ribo-

somes shift into the alternate reading frame. (tRNA^{Phe} would have three of three anticodon positions paired in the -1 frame on +3U RNA rather than two of three for tRNA^{Leu} on wild-type RNA.) Alternatively, frameshifting on the wild-type and all three +3 mutant RNAs might involve a specialized tRNA^{Leu} capable of decoding all codons with the sequence UUN (where N can be any nucleotide). To distinguish between these possibilities, we placed the +3U mutation into pGP-S (see Figure 2A) and determined the amino acid sequence at the frameshift site. The +3U-encoded transframe protein contains phenylalanine at position 11 followed by leucine at position 12 (encoded by the *pol*-frame codon UUA; not shown). Therefore, in the +3U mutant, frameshifting is mediated by a tRNA^{Phe}.

Other Retroviral Frameshift Signals Functionally Replace the RSV Signal

The sequence at the end of the simian retrovirus type 1 (SRV-1) *pro* gene exactly matches the last seven nucleotides of RSV *gag*, except that the UUA codon is substituted with UUU (Table 1), the same substitution as in the RSV +3U mutant. Given the successful substitution by the presumed SRV-1 frameshift site, we next tested two other suspected frameshift sites, A AAA AAC and U UUA AAC (Table 1), for their ability to function in place of the natural RSV frameshift sequence. As shown in Figure 4, frameshifting does occur on RNAs in which the last seven nucleotides of *gag* match these two sequences. The frameshifting efficiencies on these two RNAs are approximately 10% (Figure 4, lanes 2 and 3). Converting the last residue of these heptanucleotide sequences from C to A causes a 10-fold reduction in frameshifting efficiency (Figure 4, lanes 4 and 5). It is noteworthy that one of these mutations (lane 4) produces a run of seven consecutive A residues, yet the frameshifting efficiency is greatly reduced. This result argues that simple nucleotide redundancy is not sufficient to mediate frameshifting in this context and suggests that only certain A-site tRNAs may be competent to shift into the -1 frame. This point is strengthened by the failure of the final RSV mutant, one that replaces the RSV U UUA sequence with G GGG, to allow any detectable frameshifting (Figure 4, lane 6).

Examining the Role of RNA Secondary Structure in Frameshifting: Deletion Analysis of the RSV Stem-Loop Structure

The analysis described above proves the importance of certain nucleotides at the RSV frameshift site during the frameshifting process. However, while these nucleotides are necessary for efficient frameshifting, we suspected that they might be an insufficient signal to direct ribosomes to change frame. Based on the failure of two frameshift sites of MMTV to cause frameshifting in a novel genetic context (Jacks et al., 1987), we have proposed that potential stem-loop structures located downstream of all retroviral frameshift sites might also be required to achieve high-level frameshifting (Jacks et al., 1987, 1988). To begin to assess whether these stem-loop structures are relevant to frameshifting, we constructed a series of plasmids harboring progressive truncations of the RSV

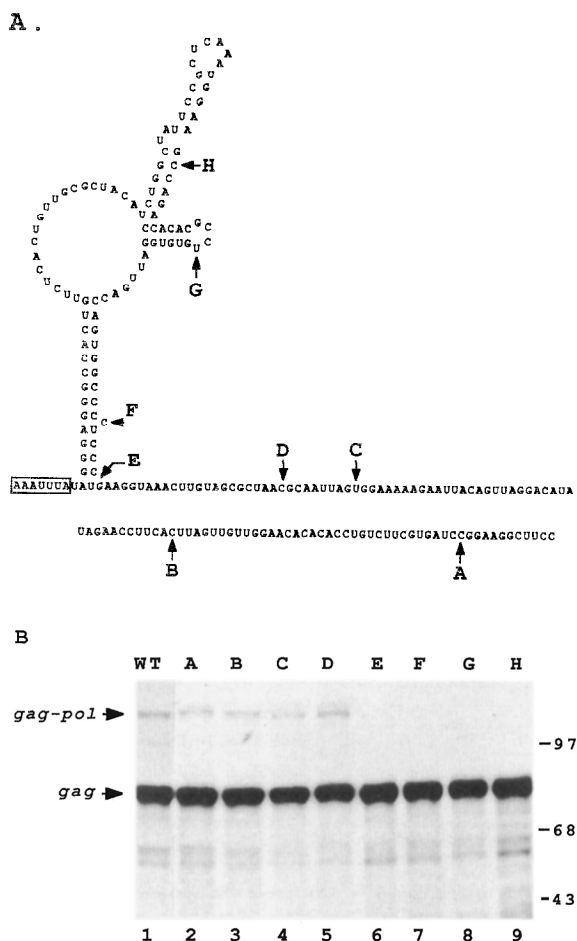


Figure 5. The Effect of Truncation of the RSV *pol* Gene on Frameshifting Efficiency

(A) The endpoints of eight truncations (A–H) in RSV *pol* are indicated by arrows relative to the sequence and predicted RNA secondary structure in the region. The RSV frameshift site is boxed. The sequence presented on the second line is a continuation of the *pol* sequence above it. In each of these mutants, the RSV sequence is followed by a constant sequence derived from the HIV-1 *pol* gene (not shown; see Experimental Procedures).

(B) Fluorogram of a 10% SDS–polyacrylamide gel containing total ³⁵S-labeled products of a rabbit reticulocyte lysate translation of RNAs transcribed from RSV truncation mutants. The names of the mutants (see [A]) are shown above the lanes. The positions of the *gag* and *gag-pol*–HIV *pol* (referred to as *gag-pol*) proteins and molecular mass markers (in kd, at right) are indicated.

pol gene. Placed downstream of the SP6 promoter, the mutant constructs contain the entire RSV *gag* gene and variable amounts of RSV *pol*, followed by a constant segment derived from the 3' end of the HIV-1 *pol* gene. (The HIV-1 segment encodes an antigen for immunoprecipitation of the protein products.) The positions of the truncation endpoints and their designations are shown with respect to a detailed model of the RSV stem–loop in Figure 5A. Frameshifting efficiency was assayed by the ability of RNA synthesized in vitro from the mutant DNAs to yield a *gag-pol* (actually *gag-pol*–HIV *pol*) fusion protein upon translation in a rabbit reticulocyte lysate system.

With one exception, those mutations that leave the predicted stem–loop structure intact give rise to wild-type levels of the *gag-pol* protein (Figure 5B, lanes 1–4). Conversely, mutations that partially or completely disrupt the structure lead to much reduced levels of frameshifting (Figure 5B, lanes 6–8). Unexpectedly, the mutant E, whose endpoint is the very last nucleotide of the predicted stem–loop structure, also shows reduced levels of frameshifting (Figure 5B, lane 5). We will discuss this result in detail below.

A 147 Nucleotide RSV Fragment Is Sufficient to Cause Frameshifting in a Novel Context

We next used these deletion mutations to determine the minimum RSV RNA sequence sufficient to allow frameshifting in a novel genetic context. We replaced all but the last 11 codons at the *gag* gene in each of the original mutants with a portion of the ground squirrel hepatitis B virus surface antigen (GS-sAg) gene such that the only RSV sequences in the resulting plasmids extend from just upstream of the frameshift site to the deletion endpoints (Figure 6A).

As shown in Figure 6B, RNAs containing the four longest RSV inserts (GS-A to GS-D) yield significant amounts of the transframe protein (the product of frameshifting) upon in vitro translation (lanes 1–4). The efficiency of frameshifting is approximately 5%, similar to that obtained with wild-type RSV RNA. The shortest fully functional RSV sequence, present in the GS-D derivative, is 147 nucleotides (Figure 6B, lane 4). It is likely that the minimum RSV sequence capable of conferring frameshifting ability is shorter than this: the first 26 nucleotides of the RSV sequences lie upstream of the frameshift site and are presumably dispensable (see Figure 3); also, the 3' boundary for sufficiency probably lies between the endpoints of the fully functional D mutant and the defective E mutant (a distance of 23 nucleotides).

As in the initial deletion analysis, the E mutation, which removes sequences up to the base of the predicted stem, leads to greatly reduced frameshifting with the GS-sAg gene segment in place of the RSV *gag* gene (Figure 6B, lane 5). The final three GS derivatives carry still fewer of the RSV stem–loop nucleotides and make even less or no transframe protein (Figure 6B, lanes 6–8).

Disrupting Base Pairs within the Stem Reduces Frameshifting Efficiency; Restoring Base Pairs Rescues It

The results presented above suggest that the RSV stem–loop is required for efficient frameshifting but show definitively only that certain sequences within *pol* are important in this event. To test directly whether the stem–loop structure itself, and not merely its primary sequence, influences frameshifting, we investigated the effects of specific stem–destabilizing and –restabilizing mutations.

Beginning with a plasmid carrying the wild-type RSV *gag* gene and a portion of the RSV *pol* gene, we constructed two site-directed mutations that each disrupt the same five consecutive base pairs (located in the center of the predicted stem) by converting to their complements

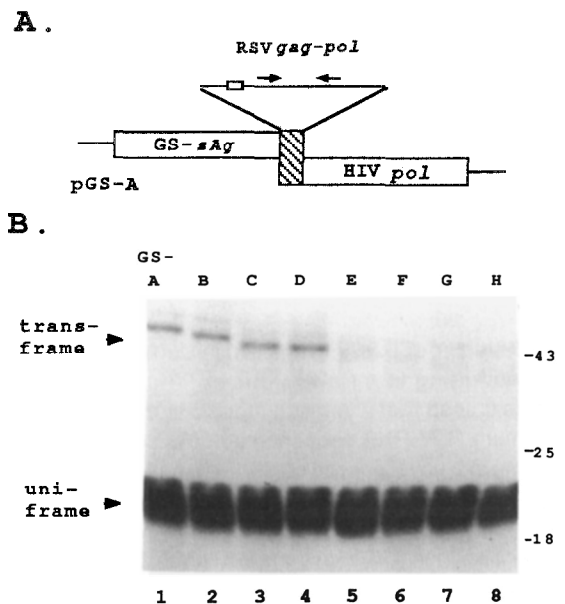


Figure 6. Frameshifting on RSV Sequences Present in a Heterologous Context

(A) A portion of the plasmid pGS-A is diagrammed. The 5' region of the GS-sAg gene was used to replace all but the final 11 codons of the RSV gag gene in the original truncation plasmid A (see Figure 5, legend). Analogous plasmids, pGS-B through -H, were also constructed. The open reading frames are drawn to indicate that frameshifting in the RSV segment is required for production of a GS-sAg-RSV gag-pol-HIV pol fusion protein. The enlargement of the RSV region shows the relative position of the frameshift site (box) and nucleotides of the predicted major stem (inverted arrows). Details of plasmid constructions are given in Experimental Procedures.

(B) Fluorogram of a 12.5% SDS-polyacrylamide gel containing anti-GS-sAg-precipitated products of a rabbit reticulocyte lysate translation of GS RNAs. The names of the corresponding RSV-truncated mutants are shown above the lanes (see Figure 5A). The positions of the uniframe protein (the N-terminal GS-sAg protein produced in the absence of frameshifting) and transframe protein (GS-sAg-RSV gag-pol-HIV pol fusion protein product of frameshifting) are shown along with the positions of molecular mass standards (in kd, at right).

the five relevant nucleotides in the 5' arm (pSM1) or 3' arm (pSM2) of the stem (Figure 7A). These mutations should severely destabilize the stem structure. We also combined the two mutations in the plasmid pSM1+2; SM1+2 RNA should form a stem-loop structure similar in thermal stability to that of wild-type RSV RNA but different from wild-type in ten nucleotide positions in the central portion of the stem.

As shown in Figure 7B, frameshifting on an RSV RNA correlates with the presence of a stem-loop structure. The frameshifting efficiency of the SM-1 and SM-2 mutants is reduced greater than 10-fold as compared with the wild-type level (Figure 7B, lanes 1-3). In contrast, when the two mutations are present together in the same RNA, restoring the potential for base pairing in the stem, the frameshift efficiency returns to approximately 2.5%, one-half the wild-type value (Figure 7B, lane 4).

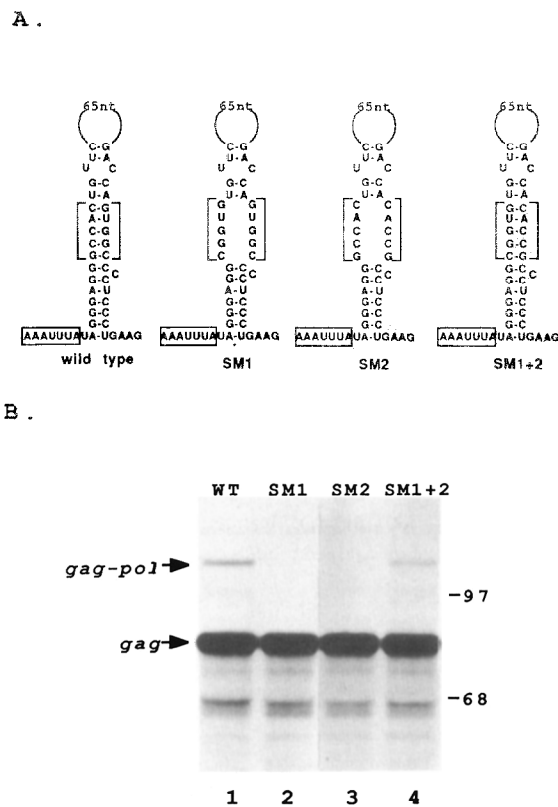


Figure 7. The Effect of Specific Stem-Destabilizing and -Restabilizing Mutations

(A) The predicted stem-loop structure of wild-type RSV RNA and three mutant derivatives. The wild-type stem-loop structure shown is a simplification of that shown in Figure 5A. The mutants SM1 and SM2 have five consecutive bases in the 5' or 3' arms of the stem changed to their complements. SM1+2 RNA carries both mutations present in SM1 and SM2 and thus can reform a stem structure. The mutants were constructed as described in Experimental Procedures.

(B) Fluorogram of a 10% SDS-polyacrylamide gel of total ³⁵S-labeled translation products of a rabbit reticulocyte lysate translation of wild-type RSV (wt), SM1, SM2, or SM1+2 RNA. The positions of the gag and gag-pol proteins and molecular mass standards (in kd, at right) are indicated.

Discussion

The discovery of ribosomal frameshifting in RSV and other retroviruses has brought to light a previously unrealized mechanism for gene expression in higher eukaryotic cells. Understanding the details of the frameshifting reaction as it occurs in retroviral gene expression may lead to the discovery of programmed frameshifts in cellular genes and should address a more general problem in translation: the accurate maintenance of reading frame.

Mechanisms of Frameshifting: Slippery Codons

Homopolymeric or "slippery" sequences have been proposed to account for frameshifting in many genes in many systems. Runs of U residues have been implicated in the -1 frameshifting during translation of gene 10 of bacteriophage T7 (Dunn and Studier, 1983) and in the +1 and -1

frameshifts inferred from the activity of leaky frameshift alleles of the yeast mitochondrial gene *oxi1* (Fox and Weiss-Brummer, 1980). The very efficient frameshift in the release factor 2 gene of *E. coli* (*RF2*) involves mispairing of the terminal 0-frame tRNA to the overlapping +1-frame codon (Craigien et al., 1985; Weiss et al., 1988). tRNA slippage by one or a few nucleotides in the 5' and 3' direction along several synthetic homopolymeric runs has recently been observed in *E. coli* by Weiss et al. (1987).

Simultaneous Slippage

The amino acid sequence at the RSV *gag-pol* frameshift site and the results of the site-directed mutagenesis presented here indicate that ribosomal frameshifting in RSV (and, by analogy, other retroviruses) is also mediated by slippage of tRNAs along homopolymeric sequences. However, the mechanism of frameshifting as it occurs in retroviral genes differs from those discussed above in that two adjacent tRNAs slip into the alternate (-1) frame. Thus, for RSV an A-site tRNA^{Leu} and P-site tRNA^{Asn} move from the last two *gag* codons into the *pol* frame, adopting a two-out-of-three base pair, anticodon-codon configuration. The requirement for at least two-of-three complementarity between the A-site tRNA and the -1-frame codon seems absolute since any change that disrupts the run of three U residues that determines this pairing abolishes frameshifting. The role of the tRNA reading the P-site codon at the RSV frameshift site, while important, is less critical. Mutations in the run of three A residues responsible for the 0 and -1 frame interactions of the tRNA^{Asn} lower the frameshifting efficiency by approximately 5-fold, but the *gag-pol* protein is still readily observed. Consistent with the more relaxed P-site requirements is the presence of several different P-site codons in various retroviral frameshift sites while only three A-site codons are observed (see Table 1). Further mutagenesis experiments are needed to define better the requirements for the P-site codon-anticodon interaction.

Mechanism of Frameshifting: Shifty tRNAs

Although the complementarity between the -1-frame codons and the anticodons of the tRNAs responsible for frameshifting on the RSV site is necessary for efficient frameshifting, such complementarity alone is insufficient to account for frameshifting in this setting. For example, given the appropriate nucleotides upstream, the sequences U UUA, U UUU, and A AAC allow frameshifting to occur when present at the end of RSV *gag*, while the sequences A AAA and G GGG are much less effective. The suggestion that only certain, specialized "shifty" tRNAs are competent to sample the alternative reading frames for suitable base-pairing interactions is also supported by the observation that in all of the documented or suspected retroviral frameshift sites (one of which is present in each of the retroviral overlaps), only three A-site codons are found: UUA, UUU, and AAC (Table 1). Two of these three (UUA and UUU) are also present as P-site codons in certain frameshift sites (Table 1). Discovery of

the special features (if any) of the tRNAs that mediate frameshifting in retroviral genes must await their purification and sequencing.

The Role of the Stem-Loop

In addition to the sequences at the frameshift site, ribosomal frameshifting at the end of the RSV *gag* gene is dependent on an RNA secondary structure located just 3' to this site. The importance of the stem-loop structure is illustrated by the inhibition of frameshifting caused by stem-disrupting mutations and, most convincingly, by the recovery of high-level frameshifting when two complementary mutations, which are separately deleterious, are combined in the same mRNA.

Possible Functions for the Stem-Loop

The presence of a downstream RNA secondary structure could influence ribosomes at the frameshift site in a number of ways. The stem or loop could be the binding site for a ribosomal protein or RNA or a soluble elongation factor; this binding could then affect the fidelity of the ribosome-tRNA interaction at the decoding sites. The stem-loop could directly displace the ribosome from its 0-frame alignment by interfering with normal tRNA translocation, effectively "pushing" the ribosome into the -1 frame. During frameshifting in the *E. coli RF2* gene, an mRNA sequence upstream of the frameshift site binds a segment of 16S RNA (that which normally binds the Shine-Dalgarno sequence during initiation). This binding is thought to facilitate movement of the ribosome into the +1 frame.

Another possible role for the stem-loop, one for which we have some experimental support, is as a translational barrier that simply causes ribosomes to pause at the frameshift site, allowing increased time for the ribosome-bound tRNAs to reach an alternative configuration on the mRNA. Translational time-course experiments with RSV RNA derivatives have shown that some fraction of ribosomes do pause at or near the frameshift site and that the extent of pausing is greatly reduced when the stem-loop structure is perturbed (T. J. and H. E. V., unpublished observations). The enhancement by adjacent stop codons of frameshifting efficiency along certain sequences in *E. coli* (Weiss et al., 1987) might also be explained by translational pausing (during decoding of the terminator), broadening the time window for tRNA realignment.

The Structure of the RNA

Throughout this text we have referred to the necessary RSV RNA secondary structure as a stem-loop, and it is clear from the effects of mutations and complementary mutations that the proposed major stem is a part of the active structure. There are indications, however, that the structure may be more complex. First, the 65 nucleotides between the two arms of the major stem are predicted to form two additional stem-loop structures (Figure 5A). Second, the E deletion described above, which leaves the proposed structure intact, nevertheless greatly inhibits frameshifting. This result is consistent with an important tertiary interaction—for example, between unpaired nucleotides in

the loop and nucleotides downstream of the major stem, a so-called pseudoknot structure (Pleij et al., 1985; Puglisi et al., 1988).

A Frameshift Cassette

The availability of a short RNA sequence capable of inducing high-level frameshifting is useful for many purposes, including the production of a fixed ratio of two proteins related at their amino termini. As shown above, all of the sequences necessary for high-level frameshifting are contained in a 147 nucleotide RSV RNA sequence. We have previously reported production of a transframe protein directed by a 50 nucleotide sequence derived from HIV-1 (Jacks et al., 1988). However, while we have observed frameshifting on these cassettes in two settings, we do not expect them to function equally well in all contexts. At least for RSV, RNA structure is critical for frameshifting, and a perturbation of that structure by new surrounding sequence would be expected to lower frameshifting efficiency. In fact, an alternative explanation for the poor efficiency of the E deletion mutant (rather than the tertiary interaction suggested above) is that the novel 3' sequence abutting the stem destabilizes the structure.

The Generality of Stem-Loop Involvement in Retroviral Frameshifting

Documentation of the importance of a stem-loop structure in the RSV frameshifting reaction supports the view that all retroviral frameshifting events are dependent on downstream RNA structure. The argument is strengthened by the presence of potential stem-loop structures (showing little primary sequence homology) downstream of all putative retroviral frameshift sites (Jacks et al., 1987, 1988, and unpublished observations). However, experiments we have recently performed with HIV-1 indicate that this view may be too simplistic. In a series of constructs in which the HIV-1 *gag-pol* sequences downstream of the frameshift site were replaced by heterologous sequences, we observed variable effects on frameshifting efficiency (Madhani et al., 1988). In one case the efficiency was reduced approximately 10-fold, while in others (including specific stem-destabilizing mutations) it was not significantly different from that determined for wild-type HIV-1 RNA (Madhani et al., 1988). These results demonstrate that sequences downstream of the HIV-1 frameshift site can influence frameshifting efficiency but also that high-level frameshifting can occur at this site, at least in vitro, in the absence of an obvious downstream stem-loop structure.

Frameshift Sites in Other Genes

Eukaryotic cells use several mechanisms to overcome the limitations of constrained translational initiation in order to express multiple protein products from individual genes. These mechanisms include: polyprotein synthesis, the production of multiple mRNAs (through the use of alternative sites of transcriptional initiation, splicing, or polyadenylation, or mRNA editing), and termination suppression. The potential for high-level ribosomal frameshifting in-

troduces yet another means to generate multiple proteins from individual genes and, in fact, individual mRNAs.

Frameshifting in eukaryotic cells is not limited to retroviruses and their related transposable elements. Brierley et al. (1987) have recently reported high-level frameshifting in the F1/F2 overlap of the coronavirus avian infectious bronchitis virus (IBV). Although the site of this -1 frameshift has not been identified, the U UUA AAC sequence contained in the F1/F2 overlap is a likely candidate. This sequence is also present in two retroviral overlaps (Table 1) and, as shown above, allows efficient frameshifting when placed at the end of RSV *gag*. The putative IBV frameshift site is also followed closely by a GC-rich stem-loop structure (Brierley et al., 1987).

To begin to investigate whether frameshifting occurs in other, nonretroviral genes, we have recently conducted a computer-assisted search of eukaryotic gene sequences for the four heptanucleotide frameshift sites shown in this report to allow efficient frameshifting in RSV *gag* (R. Colgrove, T. J., and H. E. V., unpublished). While these sequences occur much less frequently than would be expected from statistical considerations, they are found, in the correct reading frame, in many cellular and viral genes. However, only four of the potential frameshift sites uncovered in our search (all present in viral genes) are followed by stem-loop structures of significant stability. Three of these sites are present in the analogous position in three alphavirus genomes (Garoff et al., 1980; Rice and Strauss, 1981; Dalgarno et al., 1983); the fourth is located in the genome of tobacco etch virus (Allison et al., 1986). There is no independent evidence that frameshifting occurs at any of these sites. We are currently assaying them for activity in vitro.

While our search failed to identify obvious cellular candidates for frameshifting, attention to those signals that allow efficient frameshifting in retroviral genes should hasten discovery of cellular counterparts.

Experimental Procedures

Amino Acid Sequencing

GP-S RNA and +3U-S RNA were translated in 500 μ l rabbit reticulocyte translation reactions (Promega) supplemented with [³⁵S]methionine and either [³H]leucine, [³H]isoleucine, or [³H]phenylalanine (Amersham). Protein A-containing products were purified with rabbit IgG-Sepharose (Pharmacia). Subsequently, amino acid sequence analysis was performed as described (Jacks et al., 1988). The plasmids that code for the sequenced transframe proteins, pGP-S and p+3U-S, were derived from the plasmid pHSS (Jacks et al., 1988) by replacing the HIV sequences between the *AvrII* site located 6 nucleotides from the initiator AUG and the *BssHII* site that borders the protein A gene segment with an *AvrII*-*BssHII* RSV *gag-pol* fragment. (These restriction sites are located at positions 2458 and 2724 in the sequence of Schwartz et al. [1983].) The RSV fragments were isolated from the wild-type plasmid pGP (Jacks and Varmus, 1985) (pGP-S) or the +3U mutant described here (+3U-S).

Site-Directed Mutagenesis

The protocol used for site-directed mutagenesis is an adaptation of that of Lewis et al. (1983). The plasmid pGP (or mutant derivatives) was linearized at a *HpaI* site located 248 nucleotides downstream of the *gag* terminator (position 2731 in the sequence of Schwartz et al. [1983]) and briefly digested with exonuclease III (New England Biolabs). (The extent of exonuclease III digestion was assayed using mung bean

nuclease [New England Biolabs]; plasmids that had approximately 400 nucleotides removed from each end were used as the substrates for mutagenesis.) Mutagenic oligonucleotides (10 pM) were added to 0.5 µg of exonuclease III-treated plasmid in a 5 µl reaction mixture containing 50 mM Tris-HCl (pH 8.0), 20 mM KCl, 7 mM MgCl₂, 0.1 mM EDTA, and 10 mM β-mercaptoethanol, and heated to 65°C for 5 min. After cooling to room temperature (5 min), nucleotides (150 µM dCTP, 150 µM dGTP, 150 µM TTP, 50 µM dATP, and 50 µM ATP), T4 DNA ligase (0.50 U; New England Biolabs), and Klenow fragment (2.5 U; Boehringer) were added in a volume of 5 µl, and the reaction was incubated for 8–12 hr at 15°C. The reactions were then ethanol precipitated in the presence of 2.5 M ammonium acetate and used to transform *E. coli* strain HB101. Colonies harboring mutant plasmids were identified by hybridization to ³²P-labeled mutagenic oligonucleotides. The sizes of the mutagenic oligonucleotides ranged from 20–34 nucleotides. The mutations were verified by double-stranded DNA sequencing (Chen and Seeburg, 1985).

SP6 transcription reactions and rabbit reticulocyte translations were performed as described (Jacks and Varmus, 1985).

Construction of Deletion Mutants

The plasmid pGP (Jacks and Varmus, 1985) was first digested with HindIII (position 2740 in the sequence at Schwartz et al. [1983]) and treated with Bai31 (IBI) according to the specifications of the manufacturer. The ends of the DNA were then blunted with T4 DNA polymerase (Boehringer), and KpnI linkers (Collaborative Research) were added using T4 DNA ligase (IBI). After exhaustive digestion with Asp718 (an isoschizomer of KpnI) and PvuI (which cuts in the vector sequence), the resulting fragments were ligated to complementary fragments from the plasmid pAGP (Jacks et al., 1988), previously digested with Asp718 and PvuI. (The Asp718 site in pAGP is in the 3' end of the HIV-1 *pol* gene and corresponds to position 3707 in the sequence of Power et al. [1986].) The resulting plasmids were sequenced by the method of Chen and Seeburg (1985) using a primer complementary to the HIV-1 *pol* sequences. In all but two of the deletion mutants tested, the RSV *pol* and HIV *pol* sequences were in frame. For mutants B and E, the frame had to be corrected by digesting the plasmids with Asp718 and filling in the 5' overhang using the Klenow fragment of DNA polymerase I (New England Biolabs).

A second set of deletion mutants was constructed by replacing the RSV sequences upstream of the PstI site located near the end of RSV *gag* (position 2450 in Schwartz et al. [1983]) with sequences from the 5' end of the GS-sAg gene. The original truncation plasmids were cleaved with PvuI (which cuts in the vector) and PstI and were ligated to a complementary PvuI-PstI fragment from an SP6 promoter-containing plasmid (Melton et al., 1984) carrying the complete GS-sAg gene. The PstI site in GS-sAg corresponds to position 1518 in Seeger et al. (1984). SP6 transcription reactions, rabbit reticulocyte translations, and immunoprecipitations were carried out as described (Jacks and Varmus, 1985).

Acknowledgments

We thank Bruce Bowerman, Titia De Lange, Christine Guthrie, Peter Pryciak, Michael Glotzer, and Charles Craik for helpful discussion; Robin Colgrove for help with computer programs; and Janine Marinos for reluctant assistance in preparing the manuscript. T. J. is a UCSF Graduate Dean's Fellow. H. D. M. is a trainee of the Medical Scientist Training Program. H. E. V. is an American Cancer Society Professor of Molecular Virology.

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 April 12, 1988; revised August 24, 1988.

References

Allison, R., Johnston, R. E., and Dougherty, W. G. (1986). The nucleotide sequence of the coding region of tobacco etch virus genomic

RNA: evidence for synthesis of a single polypeptide. *Virology* 154, 9–20.

Brierley, I., Bournsnel, M., Birns, M., Bilmoria, B., Block, V., Brown, T., and Inglis, S. (1987). An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. *EMBO J.* 6, 3779–3785.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C., Tiollais, P., and Sonigo, P. (1987). Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. *Nature* 328, 543–547.

Chen, E. Y., and Seeburg, P. H. (1985). Supercoil sequencing: a fast and simple method for sequencing plasmid DNA. *DNA* 4, 165–170.

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

Craigie, W. J., Cook, R. G., Tate, W. P., and Caskey, C. T. (1985). Bacterial peptide chain release factors: conserved primary structure and possible frameshift regulation of release factor 2. *Proc. Natl. Acad. Sci. USA* 82, 3616–3620.

Dalgarno, L., Rice, C. M., and Strauss, J. H. (1983). Ross River virus 26S RNA: complete nucleotide sequence and deduced sequence of encoded structural proteins. *Virology* 129, 170–187.

Dunn, J. J., and Studier, F. W. (1983). Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements. *J. Mol. Biol.* 166, 477–535.

Fox, T. D., and Weiss-Brummer, B. (1980). Leaky +1 and -1 frameshift mutations at the same site in a yeast mitochondrial gene. *Nature* 288, 60–63.

Garoff, H., Frischauf, A. M., Simons, K., Lehrach, H., and Delius, H. (1980). Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. *Nature* 288, 236–241.

Guyader, M., Emerman, M., Sonigo, P., Claver, F., Montagnier, L., and Alizon, M. (1987). Genome organization and transcription of the human immunodeficiency virus type 2. *Nature* 326, 662–669.

Hiramatsu, K., Nishida, J., Naito, A., and Yoshikura, H. (1987). Molecular cloning of the closed circular provirus of human T cell leukemia virus type I: a new open reading frame in the *gag-pol* region. *J. Gen. Virol.* 68, 213–218.

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.

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

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988). Characterization of ribosomal frameshifting in HIV-1 *gag-pol* expression. *Nature* 337, 280–283.

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

Kozak, M. (1978). How do eucaryotic ribosomes select initiation regions in messenger RNA? *Cell* 15, 1109–1123.

Kozak, M. (1986). Bifunctional messenger RNAs in eukaryotes. *Cell* 47, 481–483.

Kozak, M. (1987). An analysis of 5'-noncoding regions from 699 vertebrate messenger RNAs. *Nucl. Acids Res.* 15, 8125–8148.

Lewis, E. D., Chen, S., Kumar, A., Blanck, G., Pollack, R. E., and Manley, J. L. (1983). A frameshift mutation affecting the carboxyl terminus of the simian virus 40 large tumor antigen results in a replication- and transformation-defective virus. *Proc. Natl. Acad. Sci. USA* 80, 7065–7069.

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.

- 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.
- Melitz, J. A., Grossman, Z., Luëders, K. K., and Kuff, E. L. (1987). Nucleotide sequence of a complete mouse intracisternal A-particle genome: no relationship to known aspects of particle assembly and function. *J. Virol.* 61, 3020–3029.
- 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 the yeast transposon Ty1. *Nature* 313, 243–246.
- Melton, D. A., Krieg, P. A., Rebagliati, M. R., Maniatis, T., Zinn, K., and Green, M. R. (1984). Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucl. Acids Res.* 12, 7035–7056.
- 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 are required for translation of *gag* and *pol*. *J. Virol.* 61, 480–490.
- Nilsson, B., Abrahmsen, L., and Uhlen, M. (1985). Immobilization and purification of enzymes with staphylococcal protein A gene fusion vectors. *EMBO J.* 4, 1075–1080.
- Pelham, H. R. B. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature* 272, 469–471.
- Pelham, H. R. B. (1979). Translation of tobacco rattle virus RNAs *in vitro*: four proteins from three RNAs. *Virology* 97, 256–265.
- 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.
- Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. D., Barr, P. J., and Luciw, P. A. (1986). Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. *Science* 231, 1567–1572.
- Puglisi, J. D., Wyatt, J. R., and Tinoco, I. (1988). A pseudoknotted RNA oligonucleotide. *Nature* 331, 283–286.
- Ratner, L., Hazeltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baummeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, F. (1985). Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313, 277–284.
- Rice, C. M., and Strauss, J. H. (1981). Nucleotide sequence of the 26S mRNA of sindbis virus and deduced sequence of the encoded virus structural proteins. *Proc. Natl. Acad. Sci. USA* 78, 1062–1066.
- 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.
- Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y., and Ikawa, Y. (1985). Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. *Proc. Natl. Acad. Sci. USA* 82, 677–681.
- 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.
- Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempier, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. D., Dina, D., and Luciw, P. A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227, 484–492.
- Schwartz, D. E., Tizard, R., and Gilbert, W. (1983). Nucleotide sequence of Rous sarcoma virus. *Cell* 32, 853–869.
- Seeger, C., Ganem, D., and Varmus, H. E. (1984). Nucleotide sequence of an infectious molecularly cloned genome of ground squirrel hepatitis virus. *J. Virol.* 51, 367–375.
- Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M., and Sugimura, T. (1985). Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. *Proc. Natl. Acad. Sci. USA* 82, 3101–3105.
- Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). Nucleotide sequence of Moloney murine leukemia virus. *Nature* 293, 543–548.
- Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, P., Haase, A., and Wain-Hobson, S. (1985). Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 42, 369–382.
- Sonigo, P., Barker, C., Hunter, E., and Wain-Hobson, S. (1986). Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. *Cell* 45, 375–385.
- Stephens, R. M., Casey, J. W., and Rice, N. R. (1986). Equine infectious anemia virus *gag* and *pol* genes: relatedness to visna and AIDS virus. *Science* 231, 589–594.
- Uhlen, M., Nilsson, B., Guss, B., Lindberg, M., Gatenbeck, S., and Philipson, L. (1983). Gene fusion vectors based on the gene for staphylococcal protein A. *Gene* 23, 369–378.
- Varmus, H. E. (1988). Retroviruses. *Science* 240, 1427–1435.
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* 40, 9–17.
- Weiss, R. B., Teich, N., Varmus, H. E., and Coffin, J. (1982). RNA Tumor Viruses. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).
- Weiss, R. B., Dunn, D. M., Atkins, J. F., and Gesteland, R. F. (1987). Slippery runs, shifty stops, backward steps and forward hops: -2, -1, +5 and +6 ribosomal frameshifting. *Cold Spring Harbor Symp. Quant. Biol.* 52, 687–693.
- Weiss, R. B., Dunn, D. M., Dahlberg, A. E., Atkins, J. F., and Gesteland, R. F. (1988). Reading frame switch caused by base-pair formation between the 3' end of 16S rRNA and the mRNA during elongation of protein synthesis in *Escherichia coli*. *EMBO J.* 7, 1503–1507.
- Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. J. (1985a). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. USA* 82, 1618–1622.
- Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. J. (1985b). Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. *J. Virol.* 55, 870–873.