# Expression of the Rous Sarcoma Virus *pol* Gene by Ribosomal Frameshifting

Tyler Jacks and Harold E. Varmus

Abstract. *The* pol *gene of Rous sarcoma virus is positioned downstream of the* gag *gene in a different, briefly overlapping reading frame; nevertheless, the primary translation product of* pol *is a* gag-pol *fusion protein. Two mechanisms, ribosomal frameshifting and RNA splicing, have been considered to explain this phenomenon. The frameshifting model is supported by synthesis of both* gag *protein and* gag-pol *fusion protein in a cell-free mammalian translation system programmed by a single RNA species that was synthesized from cloned viral DNA with a bacteriophage RNA polymerase. Under these conditions, the ratio of the* gag *protein to the fusion protein (about 20 to 1) is similar to that previously observed in infected cells, the frameshifting is specific for the* gag-pol *junction, and it is unaffected by large deletions in* gag. *In addition, synthesis of the fusion protein is ten times less efficient in an* Escherichia coli *cell-free translation system and cannot be explained by transcriptional errors or in vitro modification of the RNA. Ribosomal frameshifting may affect production of other proteins in higher eukaryotes, including proteins encoded by several retroviruses and transposable elements.*

Eukaryotic viruses have several strategies to generate multiple protein products from their genomes, including genome segmentation, polycistronic messenger RNA's (mRNA's), proteolytic cleavage of polyproteins, production of multiple mRNA's from a single template by differential splicing or polyadenylation, and termination suppression (*1*, *2*). Expression of the *pol* gene of several retroviruses is known to require synthesis of a polyprotein composed of regions encoded both by *pol* and by the preceding coding domain, *gag*, and later processing by proteolytic cleavage. The *gag-pol* polyprotein is produced, moreover, about 5 percent as efficiently as the translation product of *gag* alone (*3*).

The difficulties of occasionally circumventing the *gag* termination signal to make the *gag-pol* polyprotein were brought into focus by nucleotide sequencing of the genome of Rous sarcoma virus (RSV): the reading frames for *gag* and *pol* were found to differ, with *pol* in the −1 frame with respect to *gag*, and the open frames overlapping for 58 nucleotides before the *gag* amber termination codon (*4*) (Fig. 1A). Under these circumstances, synthesis of the *gag-pol* fusion protein could be achieved by either of two mechanisms: a low frequency RNA splicing event (despite the absence of conventional splice donor and acceptor sites) or occasional ribosomal frameshifting during translation, a phenomenon without precedent in higher eukaryotes (Fig. 1, B and C).

We now present strong evidence in favor of the frameshifting model, based on the synthesis of *gag* and *gag-pol* proteins in a cell-free rabbit reticulocyte translation system, when programmed with RNA transcribed in vitro from cloned RSV DNA by the bacteriophage SP6 RNA polymerase.

The rationale for our experimental approach was twofold: (i) Yoshinaka *et al.*, (*2*) demonstrated that occasional translational suppression of the amber stop codon separating the *gag* and *pol* genes of Moloney murine leukemia virus (MoMLV) is responsible for synthesis of the *gag-pol* polyprotein. Since retroviruses commonly use similar strategies for replication and gene expression (*3*), these results suggested that suppression of translational termination by ribosomal frameshifting might be used for those viruses whose *gag* and *pol* genes are out of frame. (ii) Several years ago it was reported (*5*) that RNA extracted from particles of RSV could direct the synthesis of both *gag* and *gag-pol* proteins in appropriate proportions in rabbit reticulocyte lysates (RRL). These experiments were not decisive with respect to the mechanism for generating the *gag-pol* protein, because they did not exclude the possibility that a low abundance, spliced *gag-pol* mRNA was adventitiously packaged into virions. However, if the frameshifting mechanism were correct and the spliced mRNA did not exist, the result implies that unspliced RSV RNA can be successfully translated to generate *gag-pol* polyprotein in reticulocyte lysates.

**An in vitro test for frameshifting.** To synthesize RNA in which the RSV *gag* and *pol* coding regions are in their genomic (out-of-frame) configuration, a portion of cloned RSV DNA, extending from the Sac I site 125 base pairs (bp) upstream from the *gag* initiation site to a Xba I site within *pol*, 895 bp downstream from the *gag* termination codon, was recloned in the plasmid pSP65 (*6*), in a polylinker downstream from the bacteriophage SP6 promoter (Fig. 2A). When this plasmid is linearized at the Xba I site and added to a reaction mixture containing SP6 RNA polymerase and ribonucleotide triphosphates, abundant amounts of *gag-pol* RNA (GP RNA) are synthesized. Translation of GP RNA in lysates of rabbit reticulocytes is expected to produce mainly the 76-kilodalton (kD) product of the *gag* gene, $Pr76^{gag}$. If some fraction of the ribosomes is able to shift into the −1 frame within the 58-nucleotide overlap of the *gag* and *pol* frames, a 108-kD *gag-pol* fusion protein will also be made.

Typical [$^{35}$S]methionine-labeled translation products of GP RNA are shown in Fig. 2B after fractionation by electrophoresis in sodium dodecyl sulfate (SDS)–polyacrylamide gels. Even without prior immunoprecipitation, there are clearly discernible proteins of the predicted sizes for *gag* and *gag-pol* products at the expected ratio of approximately 20:1 (Fig. 2B, lane 1). Both the 76-kD and 108-kD proteins are immunoprecipitable with serum specific for RSV $p19^{gag}$ (lane 2), but only the larger species is recognized by antiserum to reverse transcriptase, a product of *pol* (lane 3). The relatedness of the 76-kD and 108-kD proteins was further demonstrated by partial digestion with staphylococcal V8 protease. Some of the additional minor reaction products (Fig. 2B) can also be immunoprecipitated with antiserum to $p19^{gag}$ and presumably are proteins initiated at methionine codons internal to the *gag* region; for simplicity of

T. Jacks is a graduate student in the Department of Biochemistry and Biophysics and H. E. Varmus is a professor in the Departments of Microbiology and Immunology and Biochemistry and Biophysics at the University of California, San Francisco 94143.

presentation, these are not considered further, although at least one of these proteins is also precipitable by antiserum to reverse transcriptase and it presumably results from a frameshifting event similar to that which produces the 108-kD protein.

The results presented in Fig. 2B indicate that RNA containing *gag* and *pol* in different translational reading frames can direct synthesis of both *gag* and *gag-pol* proteins in vitro. Moreover, the efficiency of synthesis of the 108-kD *gag-pol* protein (about 5 percent of the level of $Pr76^{gag}$) is consistent with ratios of *gag* and *gag-pol* proteins observed in RSV-infected cells (*7*).

To begin to define the portion of GP RNA responsible for ribosomal frameshifting at 5 percent efficiency, we made the two deletion mutations of pGP shown in Fig. 3A. In GP-ΔB, we removed 463 codons of *gag* by deleting an internal Bam HI fragment from pGP. The construction maintains 51 codons at the beginning of *gag* and 189 codons at its end, so that GP-ΔB RNA can encode a 26-kD *gag* protein. If the deleted *gag* sequences are not required for frameshifting, then translation of the GP-ΔB RNA should also yield a 58-kD *gag-pol* fusion protein. Production of the predicted 58-kD *gag-pol* protein, as well as $p26^{gag}$, does occur in the absence of most of the *gag* coding domain (Fig. 3B, lane 1). The identity of the *gag-pol* protein was again confirmed by immunoprecipitation with antiserum to reverse transcriptase.

We were also able to remove a substantial portion of the 58-nt overlap region and still observe frameshifting. The plasmid pGP-ΔA was derived from pGP by deleting the 24-nt between the two Avr II sites in the overlap region (Fig. 3A). In GP-ΔA RNA, the window for frameshifting has been closed down to 34 nt. In vitro translation of this RNA produces an abundant protein that is slightly smaller than the *gag* protein of GP RNA, as was expected from the size of the deletion mutation (Fig. 3B, lanes 2 and 3). There is no apparent difference in the amounts of *gag-pol* fusion proteins produced, implying that the remaining 34 nt of the overlapping reading frames suffice to mediate frameshifting.

**Tests for RNA homogeneity**. Our conclusion that the *gag-pol* fusion protein is produced via ribosomal frameshifting relies on the assumption that the RNA synthesized and translated in vitro constitutes a homogeneous population in which *gag* and *pol* are out of frame. We have examined this assumption by addressing the unlikely possibility that a minor class of RNA, in which *gag* and *pol* occupy the same reading frame, has been generated in our experiments. This could conceivably occur either via transcriptional frameshifting by SP6 polymerase or via an unconventional type of mRNA splicing (either autogenous or induced by the in vitro conditions).

Since correction of frame could occur by the addition of one nucleotide or the deletion of two nucleotides, changes that would probably not be detectable by nuclease protection analysis, we have depended instead on a functional test for the in-frame subclass of RNA, using a cell-free translation system derived from *Escherichia coli*. To promote the efficiency of this test, we first modified the *gag-pol* DNA by the addition of a bacterial translational initiation site derived from the *E. coli* lipoprotein gene (*lpp*) (*8*); in addition, we took advantage of the proven dispensability of most of the *gag* domain (Fig. 3) and removed all of the *gag* sequences upstream of the Eco RI site preceding the *gag* terminator. The resulting construction, pLGP, and a similar plasmid, pLGP-IF, in which *gag* and *pol* are aligned in frame (Fig. 4A), were transcribed by the SP6 RNA polymerase to yield LGP and LGP-IF RNA's, which were subsequently translated in an *E. coli*–derived translation system (legend to Fig. 4).

Translation in the *E. coli* system could be directly compared to translation in the RRL because LGP RNA functions as an efficient mRNA in both systems. Moreover, the apparent efficiency of frameshifting in the RRL (approximately 15 percent) is significantly greater than observed previously. (This probably reflects a selective loss of the 10-kD *gag* protein from the polyacrylamide gel during preparation for autoradiography and not an actual increase in the level of frameshifting.) When LGP RNA is translated in both systems, similar amounts of the 10-kD *gag* protein are made, but the RRL generates at least ten times more of the 41-kD *gag-pol* protein (Fig. 4B, lanes 1 and 2). The

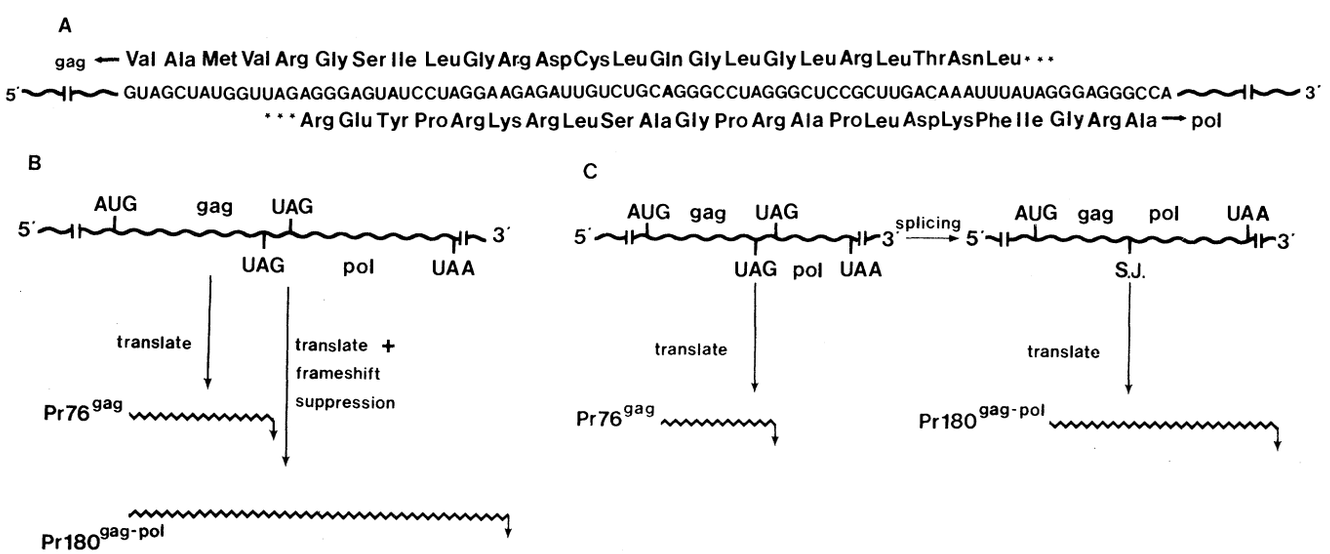


Fig. 1. Models for synthesis of the RSV proteins $Pr76^{gag}$ and $Pr180^{gag\text{-}pol}$. (A) Diagram of PR-C RSV RNA emphasizing the nucleotides shared by the *gag* and *pol* open reading frames and showing the amino acids encoded by the 3′ end of the *gag* frame and the 5′ end of the *pol* frame (*4*). The relevant stop codons are indicated by ***. (B) The translational suppression (one-mRNA) hypothesis. In this model, a single mRNA species directs synthesis of both $Pr76^{gag}$ (following normal translation) and $Pr180^{gag\text{-}pol}$ (after translational suppression of the frame difference between the *gag* and *pol* genes). (C) The splicing (two-mRNA) hypothesis. Here, infrequent splicing of the genome-length message gives rise to a second class of mRNA with a splice junction (S.J.) in which the *gag* and *pol* genes are in frame. Normal translation of these two mRNA's yields the desired proteins. (The distance between the termination codons has been exaggerated in parts B and C to emphasize the overlap region.)

reduced amount of the 41-kD fusion protein is not due to instability of the protein in the *E. coli* lysate, since translation of LGP-IF RNA yields the expected quantity of an almost identical protein (Fig. 4B, lane 3).

These results indicate that alignment of the *gag* and *pol* frames during transcription by SP6 RNA polymerase or by autocatalytic splicing cannot explain results in the RRL, since the in-frame RNA would direct equally efficient synthesis of the fusion protein in either translation system. Furthermore, the prokaryotic system appears to respond poorly to the presumptive eukaryotic signals for frameshifting at the *gag-pol* boundary.

The experiments shown in Fig. 4 do not exclude the remote possibility that synthesis of the fusion protein reflects modification (such as splicing) of the GP or LGP RNA on exposure to the RRL. However, when we retranslated RNA recovered from a RRL translation reaction in both *E. coli* system and the RRL, we observed results qualitatively similar to those obtained with newly synthesized RNA (*9*). Thus we conclude that ribosomal frameshifting during translation, rather than transcriptional frameshifting or RNA processing, is responsible for the synthesis of *gag-pol* fusion proteins in our experiments.

**Sequence specificity of the frameshifting mechanism**. It seems probable that ribosomal frameshifting in the RRL occurs in response to signals in the *gag-pol* mRNA; this would explain the differential efficiency with which the fusion proteins are made in *E. coli* and mammalian systems and would prevent promiscuous frameshifting during translation in eukaryotic cells. To test this notion more explicitly and to exclude the possibility that

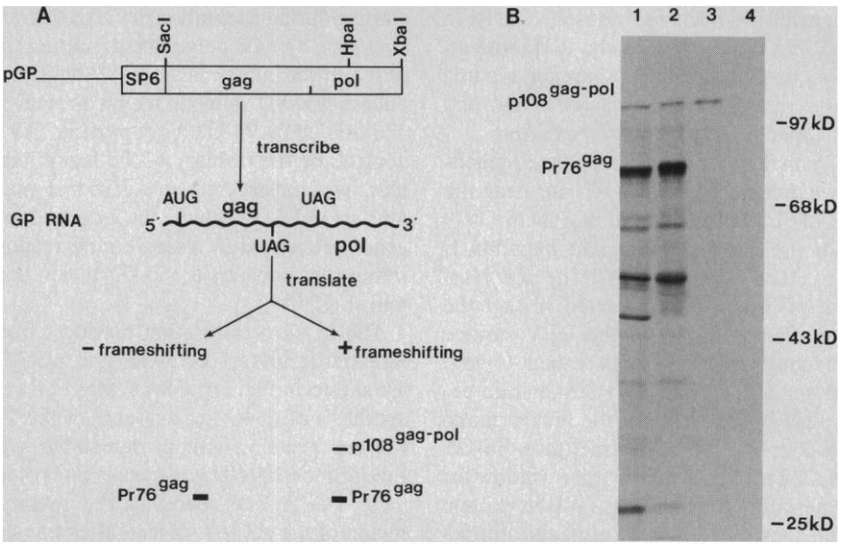


Fig. 2. An in vitro test of the translational suppression hypothesis. (A) The plasmid pGP was constructed by cloning a 3.1-kb fragment from the RSV *gag-pol* region into the SP6 vector SP65 (*6*). The RSV insert is composed of a 2.5-kb Sac I to Hpa I fragment, including the *gag-pol* overlap region, from cloned DNA of strain PR-C (*4*) and a 0.6 kb-Hpa I to Xba I fragment from strain SR-A (*31*). After linearization with Xba I, 2.5 μg of DNA were transcribed with 8 units of SP6 RNA polymerase (Promega) in a 25-μl reaction mixture (*6*). The resulting GP RNA (500 ng) was translated in a 50-μl RRL translation reaction (Promega) supplemented with [$^{35}$S]methionine (65 μCi;1200 Ci/mmol). The expected products of the translation with and without frameshifting are indicated as they would appear after gel electrophoresis. As in Fig. 1, the distance between the termination codons shown on the GP RNA is not to scale. (B) Autoradiogram of a 10 percent SDS–polyacrylamide gel of $^{35}$S-labeled products of GP RNA translation. (Lane 1) Unprecipitated (2 μl); (lane 2) immunoprecipitated (4 μl) with rabbit antiserum to p19$^{gag}$ (*32*); (lane 3) immunoprecipitated (10 μl) with rabbit anti–reverse transcriptase serum (*7*); (lane 4), immunoprecipitated (10 μl) with nonimmune rabbit serum. Immunoprecipitations were performed as described (*32*). The gel was exposed to Kodak XAR-5 film for 24 hours. Positions of labeled protein size markers are indicated.

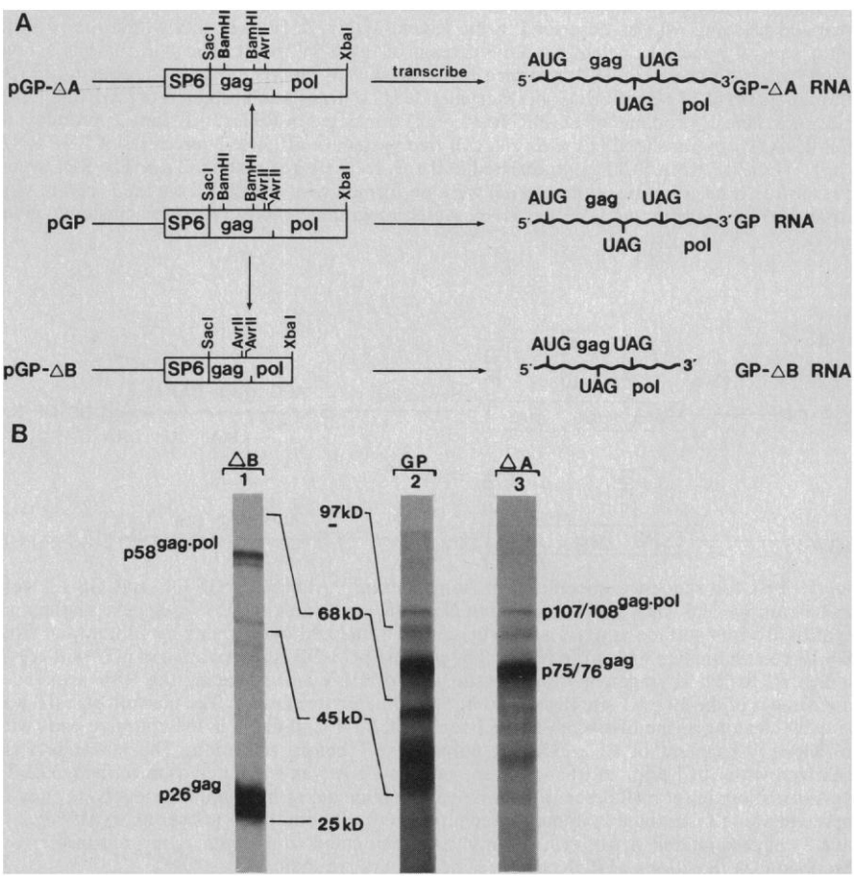


Fig. 3. Analysis of deletion mutants of pGP. (A) RNA's GP-ΔA (resulting from a 24-nt deletion in the *gag-pol* overlap region) and GP-ΔB (representing a 1386-nt internal deletion in *gag*) were transcribed from linearized plasmids pGP-ΔA and pGP-ΔB and translated (legend to Fig. 2). The plasmid pGP-ΔA was derived from pGP by cleavage with Avr II to delete 24-nt in the overlap region, followed by religation. The plasmid pGP-ΔB was constructed by cutting pGP at the two Bam HI sites within *gag*, filling in the cohesive ends with the Klenow fragment of *E. coli* DNA polymerase I (P-L Biochemicals), and religating. (B) Autoradiogram of 10 percent SDS–polyacrylamide gels of RRL translation products. (Lane 1) GP-ΔB RNA translation (3 μl); (lane 2) GP RNA translation (2 μl); (lane 3) GP-ΔA RNA translation (2 μl). Translations were performed as described (legend to Fig. 2) and products were analyzed without immunoprecipitation. Positions of molecular weight markers are indicated.

frameshifting occurred nonspecifically in the RRL, we challenged the RRL with an RNA that contains overlapping reading frames that are arbitrarily designed and not expected to induce frameshifting.

To do this, we replaced RSV sequences downstream of the Eco RI site near the end of *gag* in the plasmid pGP with DNA from the genome of human hepatitis B virus (HBV) (*10*) (Fig. 5A). The new plasmid, pGS-OF, is constructed so that the open reading frame for the HBV surface antigen ("sur") is −1 with respect to *gag*. The sur frame contains a termination codon 241 bp upstream of the first termination codon in the *gag* frame; thus GS-OF RNA offers a four times larger window for frameshifting to produce a fusion protein than does GP RNA. Nonspecific frameshifting during translation in vitro or errant synthesis by SP6 polymerase, putting the *gag* and sur genes in frame, would produce a 100-kD *gag*-sur fusion protein in addition to a 73-kD *gag* protein. As a control for the stability of the fusion protein, we constructed an additional plasmid, pGS-IF, in which the *gag* and sur genes are joined in a single long reading frame that encodes a 100-kD fusion protein (Fig. 5A).

The in vitro translation products from SP6 transcripts of pGS-OF and pGS-IF are shown in Fig. 5B. GS-OF RNA directs synthesis of abundant amounts of the 73-kD *gag* protein, without detectable synthesis of the 100-kD *gag*-sur fusion protein (lane 1). [A faint band in the relevant region of the gel is also seen after translation reactions without added RNA, and the labeled material is not immunoprecipitable with antiserum to $p19^{gag}$ (lane 2).] Translation of GS-IF RNA yields a stable 100-kD protein that is immunoprecipitable with antiserum to $p19^{gag}$ (lanes 3 and 4). Thus, within the sensitivity of these assays (about 0.1 percent), we find no evidence for unexpected frameshifting, either by transcriptional or translational mechanisms when unrelated genes are joined to produce extensive overlapping reading frames. We conclude that the production of *gag-pol* protein must depend on a mechanism that specifically recognizes some property of the *gag-pol* coding domain.

**Possible mechanisms and implications.** Our data show that ribosomal frameshift-

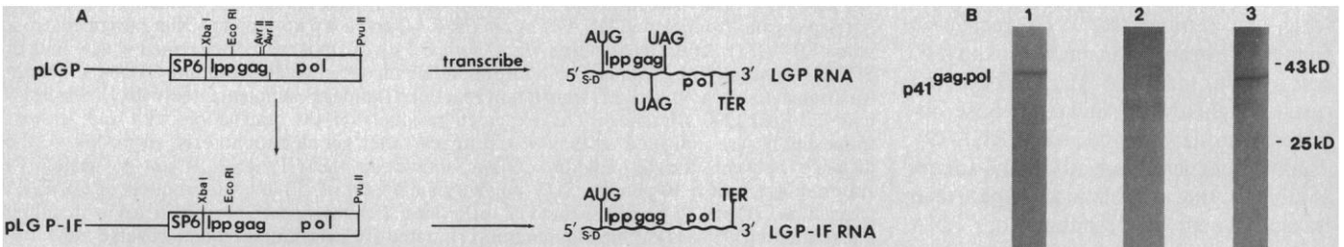


Fig. 4. Test for mRNA homogeneity. (A) Plasmid pLGP was constructed by replacing *gag* sequences upstream of the Eco RI within *gag* in pGP with sequences from the 5′ portion of the *E. coli* lipoprotein (*lpp*) gene (*8*) including the Shine-Dalgarno sequence (S-D), initiator AUG, and additional 28 codons. The plasmid pLGP-IF was derived from pLGP by cleaving with Avr II, filling in the cohesive ends with Klenow fragment, and religating. Transcription from Pvu II linearized plasmids was as described in the legend to Fig. 2. [Pvu II cuts within the plasmid vector several hundred nucleotides downstream of the RSV insert. The *pol* frame is shown closed by a stop codon (TER) presumed to be present in the vector-derived sequences.] (B) Autoradiogram of 15 percent SDS–polyacrylamide gel of translation products of LGP and LGP-IF RNA's: lane 1, products of LGP RNA (0.5 μg) translated in RRL (2 μl); lane 2, products of LGP RNA (1 μg) translated in the *E. coli* cell-free system (2 μl); lane 3, products of LGP RNA (1 μg) + LGP-IF RNA (0.15 μg) translated in the *E. coli* cell-free system (2 μl). The RRL translation was as described in legend to Fig. 2. The *E. coli* cell-free translation reactions (18 μl) were performed with reagents from an *E. coli* in vitro translation kit (Amersham) substituting an S135 extract (*34*) for the commercial S30 extract. Autoradiographic exposures were adjusted to give approximately equal intensities of the *gag* protein band.

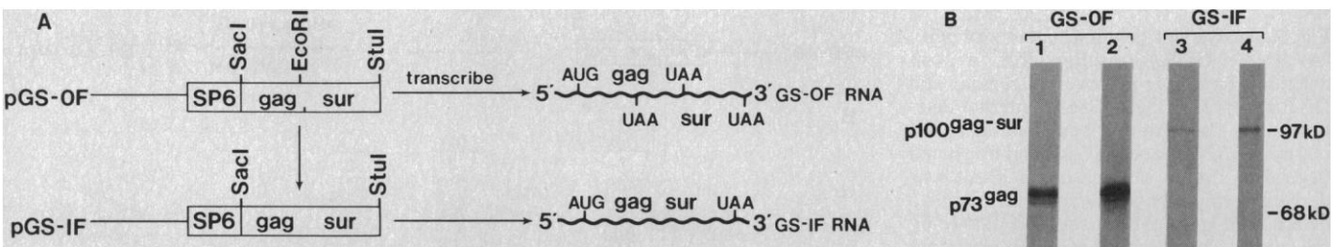


Fig. 5. Test for sequence specificity of frameshifting. (A) RNA's GS-OF and GS-IF were synthesized by SP6 RNA polymerase from pGS-OF, containing the RSV *gag* gene and human hepatitis B virus surface antigen gene (sur) out of frame and overlapping by 238 nt, and from pGS-IF containing *gag* and sur in frame. The plasmid pGS-OF is a derivative of pGP with a 1.8-kb Eco RI to Bgl II fragment from the sur gene of HBV (*10*) replacing the RSV sequences downstream of the Eco RI site that is 163 nt 5′ of the *gag* terminator. The plasmid pGS-IF was made by cleaving at the displayed Eco RI site of pGS-OF and filling in the cohesive ends with the Klenow fragment of *E. coli* DNA polymerase I before religation. The plasmids were linearized with Stu I prior to transcription, and the RNA was translated as described (Fig. 2). (B) Autoradiogram of a 10 percent SDS–polyacrylamide gel of translation products. (Lanes 1 and 2) GS-OF RNA translation products; (lane 1) unprecipitated (2 μl); (lane 2) immunoprecipitated (10 μl) with rabbit antiserum to $p19^{gag}$ (*32*); (lanes 3 and 4) GS-IF RNA translation products; (lane 3) unprecipitated (1 μl); (lane 4) immunoprecipitated (5 μl) with rabbit antiserum to $p19^{gag}$. Immunoprecipitations were performed as described (*33*). Positions of molecular weight markers are indicated.

ing occurs on an RSV RNA to produce two proteins from a single message in a mammalian cell-free translation system. The level of the *gag-pol* fusion protein produced in vitro (5 percent the level of the *gag* protein) is consistent with the ratio of these proteins in RSV-infected cells, making it likely that this form of translational suppression occurs in vivo as well. Thus, RSV may provide the first example of a ribosomal frameshift to control gene expression in a higher eukaryotic system. This mechanism has several precedents in prokaryotic systems and offers a simple way to produce a fixed ratio of two proteins. For RSV, frameshifting allows a small but predictable number of reverse transcriptase molecules to be incorporated into virus particles. Since the viral core is thought to be assembled through interactions between the genomic RNA and *gag* protein (*3*), inclusion of *pol* products in the virion reflects the ratio of *gag* to *gag-pol* proteins, which in turn, depends upon the efficiency of frameshifting. The bacteriophage T7 appears to use a very similar strategy to package a fixed ratio of two products of gene 10 into its phage heads (*11*).

Any of several mechanisms could account for the RSV frameshift.

1) Several host or viral factors might disrupt maintenance of the proper reading frame. For example, interactions between an RNA binding domain of the *gag* polyprotein and either the mRNA or a ribosomal RNA could cause ribosomes to stall during translation of the overlap region and occasionally change frame. However, the fact that the large deletions of *gag* in pGP-ΔB and in pLGP do not impair frameshifting argues against a role for *gag* protein.

2) In both bacteria and yeast, abnormal transfer RNA's (tRNA's) promote ribosomal frameshifting (*12*). These frameshift suppressor tRNA's are found in cells selected for phenotypic reversion of frameshift mutations in important genes; carry an extra nucleotide near the anticodon; and can cause +1 frameshifts at sequences related to the codons read by their normal homologs. Suppressor tRNA's of this type could not produce the RSV frameshift, which is in the −1 direction, but a role for a different type of abnormal tRNA is still possible. Any unusual tRNA would have to be present in many cell types and species though, since RSV *gag* and *gag-pol* proteins are produced in similar ratios in a wide spectrum of cells (*13*).

3) Certain normal tRNA's of *E. coli* can promote ribosomal frameshifting both in vitro (*14*) and in vivo (*15*) if present in high concentrations. These so-called "shifty" tRNA's may cause improper translocation of either two or four nucleotides resulting from a type of offset anticodon:codon pairing proposed by Weiss (*16*). Nucleotide context has been shown to influence both nonsense and missense suppression (*17*), and it almost certainly has an effect on whether a given codon will be the site of frameshifting. Short homopolymeric stretches of nucleotides have been implicated in the leakiness of frameshift mutations in the yeast mitochondrial *oxy-1* gene (*18*) and in the suspected frameshift during T7 gene 10 expression (*11*). The ribosomal frameshift that allows synthesis of *E. coli* release factor II (RFII) occurs just upstream of an amber stop codon, perhaps during a prolonged translational pause at this terminator (*19*). The RSV overlap region contains no long homopolymeric runs, but the UUA codon just 5′ of the *gag* amber stop codon is immediately preceded by another U residue (Fig. 1A). Slippage of a P-site leucine tRNA reading this UUA codon followed by mispairing to the UUU codon in the −1 frame would accomplish the necessary frameshift, and the position of the stop codon is reminiscent of the frameshift site in RFII. Placement of the frameshift at this point is also consistent with results of Rettenmier *et al.* (*20*) that suggest that the *gag-pol* fusion protein includes the arginine residue encoded in the *gag* frame just five codons upstream of the *gag* terminator (Fig. 1A). Site-directed mutagenesis of overlap nucleotides and amino acid sequencing of the fusion protein are now needed to examine these possibilities more rigorously.

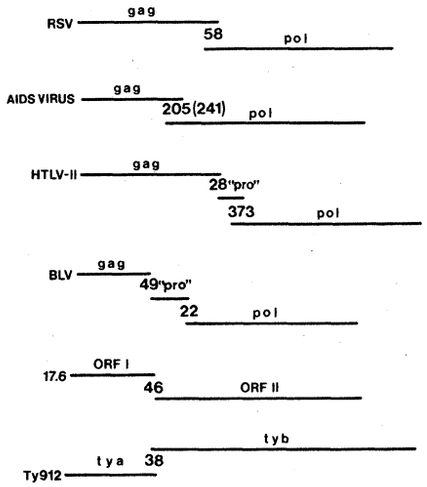


Fig. 6. Retrotransposons with overlapping of *gag* and *pol* genes. The *gag-pol* region of the AIDS retrovirus (*21*), HTLV-II (*22*), and BLV (*23, 24*), and transposable elements Ty912 (*25*) and 17.6 (*26*) are represented. Placement of one open reading frame (bar) below another signifies a reading frame in the −1 position with respect to the overlying frame. Numbers indicate nucleotides shared by the two overlapping open reading frames. (Two isolates of the AIDS virus, LAV and ARV-2, have a 205-nt *gag-pol* overlap; it is 241-nt in HTLV-III.) The protease coding domains (pro) of HTLV-II and BLV separate the *gag* and *pol* genes but overlap both. Also, the Ty912 open reading frame tyb overlaps tya in the +1 direction.

Whatever the mechanism, frameshifting in higher eukaryotes is not likely to be limited to RSV. Certain other retroviruses and some related transposable elements already appear as probable candidates (Fig. 6). The AIDS (acquired immune deficiency syndrome) retrovirus (*21*), human T-cell leukemia virus type II (HTLV-II) (*22*), and bovine leukemia virus (BLV) (*23, 24*) could all use a frameshifting mechanism to produce their (as yet unidentified) *gag-pol* fusion proteins. The latter two viruses would require two frameshift events in order to move out of the *gag* frame, through a segment encoding the viral protease, and into the *pol* frame.

The transposable elements Ty of yeast (*25*) and 17.6 of *Drosophila* (*26*), which resemble retroviruses in many other ways, also have overlapping open reading frames apparently encoding *gag*-like and *pol*-like functions. (Ty differs from the other elements by requiring a +1 shift to move from tya to tyb.) Two groups have shown that the product of the second open reading frame of Ty is expressed as a fusion protein with the product of the first, and no spliced mRNA has been detected by S1 nuclease protection analysis (*25, 27*). Ribosomal frameshifting must be considered the favored explanation for expression in these systems as well.

Is frameshifting in eukaryotes restricted to viruses and transposable elements which must subvert normal cellular machinery in order to meet their requirements for successful replication? We think this unlikely. Still, strong evidence for the cellular counterpart of the RSV frameshift is lacking. An allele of the mitochondrial *Cox* II gene, encompassing two reading frames, is conserved between two species of trypanosomes (*28, 29*), but, at least in the case of *Trypanosoma brucei*, a frame-corrected version of the gene exists and is active (*30*). The sequences of many nuclear genes include unassigned open reading frames either within or overlapping the major open reading frame. Analyses like the one described in this article might identify some of these sequences as substrates for frameshifting. Also, genetic screens could be specifically designed to uncover "shifty" sequences from eukaryotic genomes.

### References and Notes

1. B. N. Fields, Ed., *Virology* (Raven, New York, 1985); H. R. B. Pelham, *Nature (London)* **272**, 469 (1978); *Virology* **97**, 256 (1979); M. March, G. Drugeon, C. Benicourt, *ibid.* **119**, 193 (1982); E. G. Strauss, C. M. Rice, J. H. Strauss, *ibid.* **133**, 92 (1984).
2. Y. Yoshinaka *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1618 (1985).
3. R. Weiss, N. Teich, H. Varmus, J. Coffin Eds., *Molecular Biology of Tumor Viruses: RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), chaps. 4–6.
4. D. E. Schwartz, R. Tizard, W. Gilbert, *Cell* **32**, 853 (1983).
5. K. Beemon and E. Hunter, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 3302 (1977); A. Purchio, E. Erikson, R. L. Erikson, *ibid.*, p. 4661; S. R. Weiss *et al.*, *Cell* **15**, 607 (1978).
6. D. A. Melton *et al.*, *Nucleic Acids Res.* **12**, 7035 (1984).
7. H. Oppermann, J. M. Bishop, H. E. Varmus, L. Levintow, *Cell* **12**, 993 (1977).
8. J. Ghrayeb and M. Inouye, *J. Biol. Chem.* **259**, 463 (1984).
9. T. Jacks and H. E. Varmus, unpublished data. RNA was recovered from 100-μl RRL reaction mixtures with two phenol extractions and subsequent ethanol precipitation in the presence of 2.5*M* ammoniun acetate and used to program either 18-μl *E. coli* or 25-μl RRL translation reactions as in Fig. 4.
10. P. Valenzuela, M. Quiroga, J. Zaldivar, P. Gray, W. J. Rutter, in *Animal Virus Genetics*, B. Fields, R. Jaenisch, D. F. Fox, Eds. (Academic Press, New York, 1980), pp. 57–70.
11. J. J. Dunn and F. W. Studier, *J. Mol. Biol.* **166**, 477 (1983).
12. J. R. Roth, *Cell* **24**, 601 (1981).
13. V. M. Vogt, D. A. Bruckstein, A. P. Bell, *J. Virology* **44**, 725 (1982).
14. J. F. Atkins, R. F. Gesteland, B. R. Reid, C. W. Anderson, *Cell* **18**, 1119 (1979).
15. R. Weiss and J. Gallant, *Nature (London)* **302**, 389 (1983).
16. R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 5797 (1984).
17. M. Yarus and R. Thompson, in *Gene Function in Prokaryotes*, J. Beckwith, J. Davies, J. A. Gallant, Eds. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1983), pp. 23–63.
18. T. D. Fox and B. Weiss-Brummer, *Nature (London)* **288**, 60 (1980).
19. W. J. Craigen, R. G. Cook, W. P. Tate, C. T. Caskey, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3616 (1985).
20. C. W. Rettenmier, R. E. Karess, S. M. Anderson, H. Hanafusa, *J. Virology* **32**, 102 (1979).
21. S. Wain-Hobson, P. Sonigo, D. Danos, S. Cole, M. Alizon, *Cell* **40**, 9 (1985); C. A. Rattner *et al.*, *Nature (London)* **331**, 277 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 506 (1985).
22. K. Shimotohno *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 3101 (1985).
23. N. Sagata *et al.*, *ibid.*, p. 677.
24. N. R. Rice, R. M. Stevens, A. Burny, R. V. Gliden, *Virology* **142**, 357 (1985).
25. J. Clare and P. Farabaugh, *Proc. Natl. Acad. Sci. U.S.A.* **82**, 2829 (1985).
26. K. Saigo *et al.*, *Nature (London)* **312**, 659 (1985).
27. J. Mellor *et al.*, *ibid.* **313**, 243 (1985).
28. V. F. de la Cruz, W. Navelmann, L. Simpson, *J. Biol. Chem.* **259**, 15136 (1984).
29. L. Hensgens *et al.*, *Nucleic Acids Res.* **12**, 7327 (1984).
30. R. Benne, *Trends Genet.* **1**, 117 (1985).
31. W. J. Delorbe *et al.*, *J. Virology* **36**, 50 (1980).
32. H. Oppermann, J. M. Bishop, H. E. Varmus, unpublished data. Rabbit antisera were raised against gel-purified RSV $p19^{gag}$.
33. G. Ramsay, G. I. Evan, J. M. Bishop, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 7742 (1984).
34. M. Muller and G. Blobel, *ibid.*, p. 421.


35. We thank Irene Fecycz and Gunter Blobel for supplying the *E. coli* S135 extract and Pablo Garcia and Peter Walter for providing the *lpp* clone. We thank John Atkins, Bob Weiss, and Ray Gesteland for helpful discussions and materials and Janine Marinos for assistance in preparing the manuscript. Supported by a UCSF Earle C. Anthony/Dean's Fellowship (T.E.J.), American Cancer Society Research Professorship (H.E.V.), and grants from the National Institutes of Health and the American Chemical Society.

28 August 1985; accepted 30 October 1985


**RESEARCH ARTICLE**

# Detection of Single Base Substitutions by Ribonuclease Cleavage at Mismatches in RNA:DNA Duplexes


Richard M. Myers, Zoia Larin, Tom Maniatis


Physical methods for detecting single base substitutions have provided powerful tools for the analysis of human genetic diseases (*1-4*) and the establishment of human genetic linkage maps (*5–7*). These techniques could also be of considerable value in the detection and analysis of single base mutations in regulatory or protein-coding sequences. Procedures available for detecting base substitutions rely on differences in restriction endonuclease cleavage sites (*8–12*), or on differences in the melting behavior of wild-type and mutant DNA duplexes (*13–21*). For example, some single base substitutions result in the loss or gain of a restriction endonuclease cleavage site, and can therefore be detected in Southern blotting experiments (*8–12*). However, it is usually necessary to use a large number of different restriction enzymes before a change is detected. In addition, many substitutions cannot be detected by this procedure because they do not alter a restriction site. Another approach involves the use of synthetic oligodeoxyribonucleotides as differential hybridization probes (*13–16*). In this method, a labeled synthetic oligonucleotide homologous to the mutant or wild-type DNA is hybridized to blotted genomic DNA. Hybridization or washing conditions are then adjusted to allow the differential melting of the mismatched and perfectly paired duplexes. This method is useful for scoring substitutions at specific locations, but is not practical for screening large regions of DNA for new mutations or polymorphisms.

Differential DNA melting is also the basis for detecting single base substitutions by denaturing gradient gel electrophoresis (*17–21*). In this method, wild-type and mutant DNA molecules are separated by electrophoresis in polyacrylamide gels containing a gradient of formamide and urea. Duplex DNA fragments move through these gels with a constant mobility determined by molecular weight until they migrate into a portion of the gel containing a denaturant concentration sufficient to melt the DNA. When the DNA undergoes melting, its electrophoretic mobility abruptly decreases. Thus, the final position of a DNA fragment in the gel is determined by its melting temperature. The difference in melting temperature between two fragments that differ by a single base change is sufficient to allow separation on the gel. Even greater separation is achieved with DNA duplexes containing a single base mismatch (*18*). With specially designed plasmid vectors, virtually all possible single base substitutions can be detected in cloned DNA fragments (*19, 20*). However, for technical reasons (*18–21*), only 25 to 40 percent of all possible substitutions can be detected directly in total genomic DNA.

Because of the limitations in the procedures discussed above, we developed an alternative method for detecting single base substitutions in cloned and genomic DNA. This method involves the enzymatic cleavage of RNA at a single base mismatch in an RNA:DNA hybrid. The strategy used is based on the development of methods for synthesizing RNA probes (*22–24*), and on the observation that many ribonucleases are specific for single-stranded RNA under appropriate reaction conditions (*25*). A similar strategy had been developed earlier to detect mutations in duplex DNA containing single base mismatches (*26,*


Richard M. Myers is a postdoctoral fellow, Zoia Larin is a research assistant, and Tom Maniatis is a professor in the Department of Biochemistry and Molecular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138. After 1 January the address for R.M.M. is Department of Physiology, University of California, San Francisco 94143.