Translational Suppression in Gene Expression in Retroviruses and Retrotransposons*

T. JACKS

1 Introduction

The past 5 years have brought an exciting and very unexpected solution to a long-standing question in retrovirology: the mechanism of expression of the *pol* gene. Since the earliest studies of retroviral gene expression, the mechanism by which *pol*, the gene that encodes the critical enzymes reverse transcriptase, integrase, and sometimes protease, acts had remained an enigma. Experiments carried out recently seem to have finally settled this issue, as the *pol* genes of several retroviruses and one retrotransposon have been shown to be expressed by one or another form of translational suppression. This solution to the problem of *pol* gene expression is as unexpected as it is unusual. Even 5 year ago there was general agreement in this field that mRNA splicing would ultimately be found to be responsible for expression of the *pol* functions.

Whitehead Institute, Nine Cambridge Center, Cambridge, MA 02142, USA
*This work is supported by a Merck Postdoctoral Fellowship from the Helen Hay Whitney Foundation

Current Topics in Microbiology and Immunology, Vol. 157

By way of introduction to these current findings, I have chosen to review the history of this one-time enigma. This historical treatment is worthwhile both because it provides a backdrop for recent discoveries and illustrates how our preconceptions about the way things work can sometimes lead us astray.

pol is the central of the three replication genes carried by all replication-competent retroviruses (WEISS et al. 1984). It is preceded by the *gag* gene (encoding the structural genes of the virus core) and followed by the *env* gene (encoding the glycoproteins of the viral membrane) (WEISS et al. 1984). Examination of the mRNAs encoded by retroviral proviruses in the mid-1970s immediately suggested how *gag* and *env* were expressed. Two major messages were found in infected cells: one (the genome-length message) carried *gag* at its 5'-terminus, and another (the spliced subgenomic mRNA) with *env* in the 5' proximal position (HAYWARD 1977; WEISS et al. 1977). Following the general rule in eukaryotic cells that limits translation to the 5'-most open reading frame in a given mRNA (KOZAK 1978), these two messages should (and do) encode the Gag and Env proteins (VON DER HELM and DUESBERG 1975; PAWSON et al. 1976; PURCHIO et al. 1977; KERR et al. 1976; MURPHY et al. 1979; STACEY et al. 1977). What then is the mRNA for the *pol* gene?

The solution to what might have been called the "*pol* problem" came from the analysis of *pol*-encoded proteins in virus-infected cells. Using an antiserum specific for the *pol* product reverse transcriptase and another directed against a Gag antigen, OPPERMAN et al. (1977) showed that the primary translation product of the *pol* gene of RSV was, in fact, a Gag-Pol fusion protein. Pulse-chase experiments showed that this fusion protein was not the precursor to the Gag protein, which was present in infected cells approximately 20-fold more abundantly than the Gag-Pol protein (OPPERMANN et al. 1977). Moreover, tryptic peptide analysis of the Gag-Pol fusion proteins of RSV and MLVs indicated that they contained most, if not all, of the sequences present in the respective Gag proteins (OPPERMAN et al. 1977; JAMJOON et al. 1977; RETTENMEIR et al. 1979). Thus, the *pol* problem was transformed into the perhaps more interesting *gag-pol* problem: how could an apparently single species of mRNA (the genome-length mRNA) give rise to both the Gag and Gag-Pol proteins?

At the time the *gag-pol* problem was defined, two hypotheses were advanced as potential solutions. According to the suppression hypothesis (Fig. 1), the genome-length mRNA is translated to yield both the Gag and Gag-Pol proteins, with the latter arising upon occasional suppression of the signal(s) that normally terminates translation at the end of the *gag* gene. The splicing hypothesis, on the other hand, calls for the inefficient processing of the genome-length message to generate a rare species of "*gag-pol*" mRNA in which the two genes are joined in one long open reading frame (Fig. 1).

The first attempt to discern the actual mechanism of *pol* gene expression came in 1978 when PHILIPSON and coworkers (1978) translated MLV virion RNA (vRNA) in an in vitro translation system supplemented with yeast suppressor tRNAs. Several groups had previously shown that cell-free translation of MLV or RSV vRNA (or purified genome-length mRNA) yielded gag and gag-pol proteins

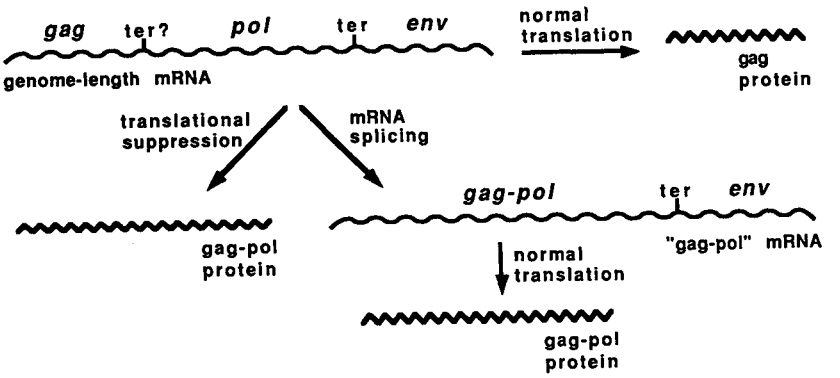


Fig. 1. Models for *pol* gene expression. Normal translation of the retroviral genome-length mRNA is shown to give rise to Gag protein. Generation of the Gag-Pol protein from this mRNA could be accomplished by either translational suppression of the signal(s) that normally terminates translation at the end of *gag* and prevents translation into *pol* or production of a specific "*gag-pol*" mRNA (in which the coding domains of *gag* and *pol* are fused into one long open reading frame) through mRNA splicing

in ratios similar to those observed in infected cells (VON DER HELM and DUESBERG 1975; PAWSON et al. 1976; PURCHIO et al. 1977; KERR et al. 1976; MURPHY et al. 1979).[1] PHILIPSON et al. (1978) noted that addition of yeast amber suppressor tRNA to an MLV-vRNA-programmed rabbit reticulocyte lysate translation reaction enhanced production of the Gag-Pol protein at the expense of the Gag protein. This observation strongly suggested that the MLV *gag* and *pol* genes were in the same translational reading frame and separated by a single amber termination codon, a configuration that was at least consistent with the suppression hypothesis. This presumed *gag-pol* configuration was later confirmed by DNA sequencing of an MLV provirus (SHINNICK et al. 1981), but neither the sequence nor the in vitro suppression of the *gag* terminator guranteed that translational suppression was the actual mechanism of MLV *gag-pol* expression in vivo.

In fact, shortly after the report that the MLV Gag-Pol protein could be synthesized in vitro by the addition of nonsense suppressor tRNAs, a similar experiment performed with RSV vRNA produced a contrary result. As had been found with MLV, WEISS et al. (1978) observed that in vitro translation of RSV vRNA in the presence of yeast amber suppressor tRNA reduced the yield of Gag protein. However, rather than producing a corresponding increase in the level of the Gag-Pol protein, this treatment resulted in appearance of a novel, extended Gag protein and no additional Gag-Pol protein. The conclusion from this experiment was that the RSV *gag* gene is terminated by an amber stop codon, but

[1] Note that the fact that vRNA can be translated to yield both Gag and Gag-Pol proteins does not by itself distinguish between the suppression and splicing hypotheses since virions could contain both the genome-length mRNA and the potentially very similar, spliced *gag-pol* mRNA

unlike the situation with MLV this terminator is not immediately followed by an inframe *pol* gene. At least one more terminator or a difference in reading frame WEISS and co-authors reasoned stands between the RSV *gag* and *pol* genes. Since they doubted that multiple stop codon or frameshift suppression would be adequately efficient to yield the observed ratio of Gag to Gag-Pol proteins, these authors argued that the most likely mode of *gag-pol* expression of RSV was via the production of a spliced *gag-pol* mRNA (see Fig. 1).

The nucleotide sequence of RSV reported by SCHWARTZ et al. (1983) clarified the genetic structure of the RSV *gag-pol* region. The conclusions drawn from this sequence, however, might have further delayed the ultimate solution to the *gag-pol* problem. Consistent with the in vitro translation data, SCHWARTZ et al. (1983) found that the RSV *gag* gene terminates with the amber stop codon and that this stop codon is followed by a second one in the *gag* reading frame some 111 nucleotides downstream. The *pol* open reading frame (identified by its position relative to *gag* and the presence of a coding region whose predicted amino acid sequence matched the known N-terminal acid sequence of RSV reserse transcriptase) is in a different translational reading frame than *gag*. The 5'-end of the *pol* open reading frame overlaps the 3'-end of *gag* by 58 nucleotides in the -1 direction. As defined by SCHWARTZ et al. (1983) however, the 5'-end of the *pol* "gene" begins with the portion known to encode reverse transcriptase, located 20 nucleotides downstream of the *gag* terminator. While acknowledging the possibility that ribosomes could shift reading frame during translation of the 58 nucleotide *gag-pol* overlap, these workers firmly concluded that the only reasonable way to synthesize the RSV Gag-Pol protein would be form an RNA derived from the genome-length mRNA by splicing that carried the *gag* and *pol* genes fused in-frame.

The view, first formed with respect to RSV, soon dominated the field of retrovirology generally. This bias is indicated most obviously in the treatment of the subject in the comprehensive review, *RNA Tumor Viruses* (WEISS et al. 1984). Largely based on the evidence presented above, the authors of several chapters allude to the near necessity for an RSV *gag-pol* mRNA. At one point it is claimed that such a species "must" exist (p. 581). Since the replication strategies of the different retroviruses are similar, it was also generally believed that MLV, for which stop codon suppression was at least a structural possibility, and other retroviruses also expressed the Gag-Pol protein from a separate, spliced mRNA.

The splicing hypothesis gained more credibility with the first nucleotide sequence of human T-cell leukemia virus type I (HTLV-I) SEIKI et al. 1983). This sequence included a 300 nucleotide "intergenic" region between *gag* and *pol* that was closed in all three reading frames by multiple termination codons. As such, translation from *gag* into *pol* along the genome-length mRNA would require multiple suppression events. Although physical evidence for a spliced *gag-pol* mRNA was lacking for this or any other virus, this presumed *gag-pol* "intron" seemed to leave no alternative.

In the year 1983, then, the *gag-pol* problem seemed ostensibly solved, awaiting only the isolation and characterization of the elusive *gag-pol* mRNA. In

the intervening 6 years, however, the consensus opinion has taken an about-face: the weight of the available evidence favors the view that translational suppression accounts for Gag-Pol synthesis in all retroviruses and many retrotransposons. In the body of this review, I will discuss the experiments that led to the transformation of opinion away from the splicing hypothesis and toward translational suppression. But first, I will conclude this introduction by briefly considering how the incorrect solution became so popular.

The general acceptance of the splicing hypothesis occurred primarily on account of the interpretation of the data concerning RSV (WEISS et al. 1977; SCHWARTZ et al. 1983) and the seemingly irrefutable evidence from HTLV-1 (SEIKI et al. 1983). These papers are, in fact, often cited as evidence for splicing in retroviral *pol* gene expression. With hindsight one can now suggest that both WEISS et al. (1977) and SCHWARTZ et al. (1983) should have been more even-handed with regard to the possible mechanisams of *pol* gene expression. However, they connot be faulted for favoring a mechanism that was fast becoming the norm in eukaryotic gene expression (mRNA splicing) over one that had no physiological precedent (frameshift suppression). And while these publications strongly influenced the field's perception of the *gag-pol* problem, they actually only solidified an existent bias found generally in favor of mRNA splicing.

More importantly, the conclusions drawn from the RSV data set the stage for what seemed at the time to be overwhelming evidence for splicing in HTLV-1 (SEIKI et al. 1983). In this case there appeared to be no need for interpretation. And, indeed, even now, if faced with a genetic wasteland between the *gag* and *pol* genes of a given virus, one would have to conclude that splicing would be required to generate a joint Gag-Pol protein for that virus. But, in fact, no such retrovirus is known to exist. The originally sequenced clone of HTLV-1 is almost certainly noninfectious. Based on the sequences of two additional HTLV-1 clones (HIRAMATSU et al. 1987; NAM and HATANAKA 1986), it is now clear that the region between HTLV-1 *gag* and *pol* is a coding domain. The region comprises an open reading frame whose predicted product is homologous to known retroviral proteases. Furthermore, the so-called *pro* gene overlaps the 3'-end of *gag* and the 5'-end of *pol*. This overlapping, three gene structure has been observed in several other retroviruses as well (Fig. 2).

Thus, the death knell for the suppression hypothesis was sounded on account of the sequence of a noninfectious clone. This sequence and the interpretations of it, both by the authors and the retrovirological community at large, clearly illustrate the danger of preconceptions. While SEIKI et al. (1983) acknowledge that their clone was not known to be infectious, they nonetheless drew conclusions about HTLV-1 replication based on it. Given that splicing was the accepted mechanism of RSV *gag-pol* expression, the discovery of a putative intron between HTLV-1 *gag* and *pol* did not signal that something might be amiss. It should be noted that a second group later "confirmed" the presence of an intergenic region between *gag* and *pol* of HTLV-1 by sequencing a second, noninfectious HTLV-1 provirus (RATNER et al. 1985b). These are not example of making the data fit the

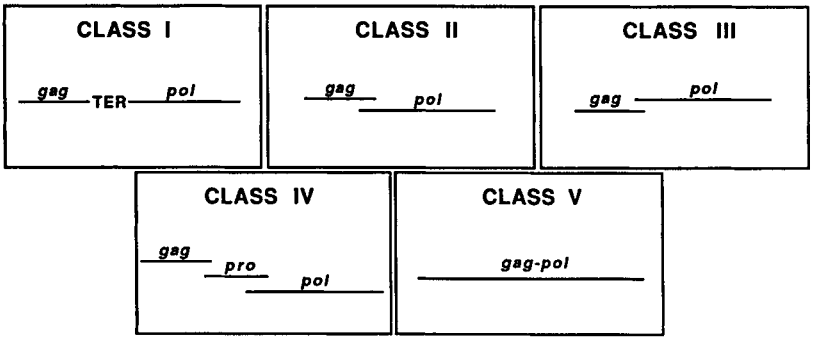


Fig. 2. The genetic structure of the *gag-pol* domains of retroviruses and retrotransposons. *Class I*: *gag* and *pol* in the same translational reading frame separated by a single termination codon (*TER*) MLV (SHINNICK et al. 1981), FeLV (YOSHINAKA et al. 1985b), and baboon endogenous virus (TAMURA et al. 1983) exhibit this arrangement. Class II: *pol* directly overlapping *gag* in the -1 reading frame. Examples of this type include: RSV (SCHWARTZ et al. 1983), HIV-1 (WAIN-HOBSON et al. 1985; RATNER et al. 1985a, SANCHEZ-PESCADOR et al. 1985), HIV-2 (GUYADER et al. 1987), simian immunodeficiency virus (CHAKRABARTI et al. 1987), Visna virus (SONIGO et al. 1985), equine infectious anemia virus (STEPHENS et al. 1986), mouse intracisternal A particle (MEITZ et al. 1987), 17.6 (SAIGO et al. 1985), and gypsy (MARLOR et al. 1987). Class III: *pol* directly overlapping *gag* in the +1 direction. The yeast transposable elements Tyl (CLARE and FARABAUGH 1985; MELLOR et al. 1985) and Ty2 (WILSON et al. 1986) and the murine element L1Md (LOEB et al. 1986) are in this class. Class IV: *gag* and *pol* separated by a third gene (*pro* encoding the viral protease) that overlaps them both. The *pro* and *pol* genes lie in the −1 frame relative to the genes that precede them (*gag* and *pro*). Retroviruses in the class IV category include: MMTV (JACKS et al. 1987; MOORE et al. 1987), simian retrovirus type 1 (POWER et al. 1986), Mason-Pfizer monkey virus (SONIGO et al. 1986), bovine leukemia viurs (SAGATA et al. 1985; RICE et al. 1985), HTLV-1 (NAM and HATANAKA 1986; HIRAMATSU et al. 1987), and HTLV-2 (SHIMOTOHNO et al. 1985). Class V: *gag* and *pol* domains contained in one long open reading frame. Three retrotransposons, *copia* (MOUNT and RUBIN 1985), *Tal* (VAYTAS and AUSUBEL 1988), and *Tnt1* (GRANDBASTIEN et al. 1989), belong to this class

theory; the sequencing data are presumably accurate. Yet the overinterpretation of the data, in light of the uncertainty about the clones, was clearly fitted to the prevailing theory and, as is discussed from here on, an incorrect theory at that.

2 Retrovirus and Retrotransposon Nucleotide Sequences and *gag-pol* Structures

The past 6 years have brought an explosion of nucleotide sequence 5 of retroviruses and retrotransposons from species ranging from yeast to humans. The genetic structures in the *gag-pol* regions of these elements fall into five classes. As shown in Fig. 2, the first class of elements carry *gag* and *pol* in the same translational reading frame separated by a single amber termination codon (Class I). In Class II elements, the 5′-end of the *pol* open reading frame overlaps the 3′-end of *gag*, with the *pol* frame offset by one nucleotide in the 5′ direction (−1) with respect to the *gag* frame. Class III elements also display directly

overlapping *gag* and *pol* genes, but for these the *pol* frame is + 1 relative to the *gag* frame. Six known retroviruses carry an additional open reading frame between *gag* and *pol* that overlaps them both (Class IV). This open reading frame encodes the viral protease and is termed variously "*prt*" and "*pro*"; *pro* is in the − 1 frame relative to *gag* and in the + 1 frame relative to *pol*. Finally, three retrotransposons appear to include both the *gag* and *pol* coding domains in one continuous open reading frame (Class V). For those elements that have overlapping genes, the size of the overlap (defined as the sequence shared by the two open reading frames) ranges from 14 to 205 nucleotides.

3 Discovery of Translational Suppression During *gag-pol* Synthesis

Members of each of the first four classes of viruses and transposons shown in Fig. 2 are currently known or believed to utilize translational suppression in the synthesis of their Gag fusion proteins. The experiments that led to these conclusions will be described in turn below. As for the retrotransposons designated Class V, in which *gag* and *pol* coding domains share the same long open reading frame, translational suppression is not required for *gag-pol* expression. The mechanism of *gag* expression in these elements is discussed in Sect. 6.

3.1 Class I: Termination Suppression

The first compelling evidence in favor of translational suppression during Gag-Pol synthesis for any retrovirus or retrotransposon came from amino acid sequence analysis of the protease protein of MLV in 1985 (YOSHINAKA et al. 1985a). This protein is initially expressed as part of the MLV Gag-Pol protein; it is responsible for cleaving itself and other mature viral proteins from their precursors. Crude mapping and sequence comparisons had suggested that the MLV protease was encoded upstream of the reverse transcriptase domain, near the 5'-end of *pol* (LEVIN et al. 1984). The N-terminal sequence of the purified protease produced by YOSHINAKA et al. (1985a) revealed that in fact the protein is encoded across the *gag-pol* junction. The first four amino acids of the protease are encoded by the last four codons of *gag*; the fifth amino acid is a glutamine; and the remainder of the protein is encoded by *pol*, beginning with the codon that immediately follows the *gag* terminator. From this amino acid sequence, it was simple to deduce the mechanism of MLV *gag-pol* expression: suppression of the *gag* amber termination codon by a glutamine-charged tRNA. Since all of the nucleotides at the *gag-pol* junction were required to encode the protease, a spliced *gag-pol* mRNA was definitively excluded.

YOSHINAKA et al. (1985b) subsequently sequenced the protease protein of FeLV, another member of the Class I elements shown in Fig. 2. The amino acid sequence once again revealed that the FeLV Gag-Pol protein is expressed via insertion of a glutamine residue in response to the *gag* amber terminator.

Termination suppression at the end of the MLV *gag* gene is not restricted to a UAG terminator. FENG et al. (1989b) have recently reported efficient suppression of the two other termination codons, UAA and UGA, when placed at the end of MLV *gag*. Proviruses harboring either UAA or UGA terminators yielded wild-type levels of virus and *pol* gene products after transfection into tissue culture cells, and in vitro translation of mRNAs transcribed from these mutants produced the normal ratio (1:20) of Gag-Pol to Gag proteins. The fact that all three termination codons are efficiently suppressed at this site suggests that features of the surrounding sequence influence the suppression event (see Sect. 4.2). It is not known which amino acids are inserted in response to the UAA and UGA codons in this setting.

3.2 Class II: −1 Ribosomal Frameshifting

Once termination suppression had been demonstrated for MLV Gag-Pol synthesis, attention quickly turned to retroviruses and retrotransposons whose *gag* and *pol* genes overlapped. If the basic replication strategies of different retroviruses are similar, these viruses should utilize another form of translational suppression, frameshift suppression, in the synthesis of their Gag fusion proteins. JACKS and VARMUS (1985) tested this possibility for RSV by cloning a DNA fragment derived from the *gag-pol* domain downstream of the *Salmonella* phage 6 (SP6) promoter. In vitro transcription of this clone by SP6 RNA polymerase yielded a homogeneous population of mRNA that mimicked, at least in the *gag-pol* region, the RSV genome-length mRNA. Translation of this synthetic mRNA in a rabbit reticulocyte lysate translation reaction would be expected to generate the Gag polyprotein. However, if some fraction of the ribosomes were able to shift into the -1 reading frame during translation of the 58 nucleotide *gag-pol* overlap, the Gag-Pol protein would also be produced. The result was clear-cut: Both Gag and Gag-Pol proteins were observed, and their ratio (approximately 20:1) closely matched that observed in RSV-infected cells (OPPERMAN et al. 1977). After excluding transcriptional frameshifting and in vitro splicing of the SP6-produced mRNA, these authors concluded that the RSV Gag-Pol protein could be synthesized in vitro from the genome-length mRNA via ribosomal frameshifting. Moreover, the efficiency of frameshifting observed in vitro ($\sim 5\%$) was sufficient to suggest that frameshifting was the mechanism of RSV *gag-pol* expression in vivo as well.

The same experimental strategy was later used to ascertain whether ribosomal frameshifting was responsible for *gag-pol* expression of human immunodefiency virus type 1 (HIV-1) (JACKS et al. 1988a), another of the Class II elements (Fig. 2). Just as with RSV, in vitro translation of a synthetic mRNA

Table 1. Heptanucleotide frameshift sites. Common 7-nucleotide sequence motifs are present in all retroviral and retrotransposon overlaps known or presumed to contain sites of frameshifting. The heptanucleotides are shown (*upper case*) along with their neighboring sequences and their distance (in nucleotides) upstream of the 0-frame termination codon. *Triplets* denote codons in the 0-frame. References for nucleotide sequences are found in the legend to Fig. 2. Table is adapted from JACKS et al. 1988b)

Retrovirus or retrotransposon	Overlap	Sequence	Distance upstream of 0-frame terminator
RSV	gag/pol	ACA AAU UUA UAG	0
HIV-1	gag/pol	AAU UUU UUA GGG	198
HIV-2	gag/pol	GGU UUU UUA GGA	267
SIV	gag/pol	GGU UUU UUA GGC	213
Gypsy	gag/pol	AAU UUU UUA GGG	51
MMTV	pro/pol	CAG GAU UUA UGA	0
SRV-1	pro/pol	GGA AAU UUU UAA	0
MPMV	pro/pol	GGA AAU UUU UAA	0
17.6	gag/pol	GAA AAU UUU CAG	30
Mouse IAP	gag/pol	CUG GGU UUU CCU	3
MMTV	gag/pro	UCA AAA AAC UUG	3
BLV	gag/pro	UCA AAA AAC UAA	0
HTLV-1	gag/pro	CCA AAA AAC UCC	18
HTLV-2	gag/pro	GGA AAA AAC UCC	18
EIAV	gag/pol	CCA AAA AAC GGG	195
BLV	pro/pol	CCU UUA AAC UAG	0
HTLV-1	pro/pol	CCU UUA AAC CAG	156
HTLV-2	pro/pol	CCU UUA AAC CUG	18
SRV-I	gag/pro	CAG GGA AAC GAC	147
MPMV	gag/pro	CAG GGA AAC GGG	147
Visna	gag/pol	CAG GGA AAC AAC	45

carrying the HIV-1 *gag* and *pol* genes in their genomic, out-of-frame configuration yielded both Gag and Gag-Pol proteins. The Gag to Gag-Pol protein ratio observed in vitro for HIV-1 was approximately 10:1, suggesting that frameshifting is more efficient on the HIV-1 mRNA than on the RSV message. It is not known whether this higher efficiency also occurs in vivo, since there are not yet accurate estimates of the ratio of Gag to Gag-Pol proteins in HIV-1-infected cells.

While the other retroviruses belonging to this class have not been directly tested, they are likely to utilize ribosomal frameshifting also. Putative frameshift signals are present in the *gag-pol* overlaps of each of them (see Table 1).

3.3 Class III: + 1 Ribosomal Frameshifting

In two known retrotransposons, the *pol* reading frame is offset by one nucleotide in the 3′ direction relative to *gag*. Thus, translation from *gag* into *pol* would require + 1 ribosomal frameshifting for these elements. The *gag-pol* expression in

one member of this class, TY of *Saccharomyces cerevisiae*, has been examined by the laboratories of Farabough and Kingsman and Kingsman. Both groups have monitored expression of the *pol*-like gene (*tyb*) by inserting a reporter gene (lacZ or α-interferon) just downstream of the *tya(gag)-tyb* overlap. Despite the absence of an initiator methionine codon in the *pol* frame upstream of the reporter genes, a high level of expression of these genes was observed (CLARE and FARABOUGH 1985; MELLOR et al. 1985). Furthermore, Western blot analysis detected the reporter proteins at a molecular weight consistent with them being fused to the product of the upstream *tya* gene. Synthesis of this fusion protein did not appear to require mRNA splicing, as Northern blot and S1 nuclease analysis failed to detect a spliced mRNA species. As had been true in the earlier studies characterizing retroviral mRNAs, this type of analysis cannot exclude a very small splice in the *gag-pol* overlap. However, it is suggestive that Ty-1 utilizes ribosomal frameshifting in the expression of its Gag-Pol protein.

This claim has been strengthened by further experiments performed by the same two groups. WILSON et al. (1986) reported that the more sensitive S1 analysis, capable of detecting a splice as small as five nucleotides, still failed to detect a spliced *tya-tyb* mRNA in yeast cells. CLARE et al. (1988) used direct mRNA sequencing to rule out both splicing and mRNA editing of the *tya-tyb* mRNA. Thus, production of the Tya-Tyb fusion protein is a posttranscriptional event, almost certainly +1 ribosomal frameshifting. CLARE et al. (1988) also noted that the β-galactosidase activity in a yeast strain containing a frameshift-requiring lacZ fusion was only fivefold below that obtained with an "in-frame" control, indicating a frameshifting efficiency of approximately 20%.

3.4 Class IV: Double Ribosomal Frameshifting

The retroviruses in this class represent the greatest challenge to the suppression hypothesis. In order to continue translation from *gag* to *pol* on the genome-length mRNAs of these viruses, ribosomes would have to change reading frames twice, first during translation of the *gag-pro* overlap and then again in the *pro-pol* overlap (Fig. 2). (Both of these frameshifts would be in the −1 direction.) Ribosomes that only shifted frame in the *gag-pro* overlap would be expected to generate a Gag-Pro fusion protein. Such a fusion has been observed in cells infected by MMTV (DICKSON and ATTERWILL 1979). Furthermore, if the ratio of the Gag to Gag-Pro-Pol proteins of viruses in this class is similar to the Gag to Gag-Pol protein ratios seen in RSV- and MLV-infected cells (and this appears to be true for, at least, MMTV), the efficiency of frameshifting in at least one of the two overlaps would need to be higher than the 5%–10% previously observed for RSV (JACKS and VARMUS 1985) and HIV-1 (JACKS et al. 1988a). (Two successive frameshift events of 10% efficiency would result in a Gag to Gag-Pro-Pol protein ratio of 100 to 1.)

Ribosomal frameshifting on MMTV mRNA has been examined in two ways. MOORE et al. (1987) and JACKS et al. (1987) used the same in vitro assay for

frameshifting as discussed above. They synthesized artificial mRNAs in vitro containing the *gag-pro-pol* portion of the MMTV genome. In addition to yielding the Gag protein, in vitro translation of these messages produced a Gag-Pro fusion protein and to a lesser extent a Gag-Pro-Pol fusion, the two expected products of frameshifting. The identities of the products were confirmed by immunoprecipitation and by truncation of the DNA templates at numerous positions prior to transcription. The comparative yield of the Gag-specific protein and the two Gag fusions indicated frameshifting efficiencies in the *gag-pro* and *pro-pol* overlaps of approximately 25% and 10%, respectively. Thus, one in four translating ribosomes changes frame in the *gag-pro* overlap, and of those, one in ten shifts into the *pol* frame in the *pro-pol* overlap. The ratio of the presumed Gag, Gag-Pro, and Gag-Pro-Pol proteins seen in MMTV-infected cells is approximately 30:10:1 (DICKSON and ATTERWILL 1979). Once again, the frameshifting efficiencies derived in vitro are consistent with this being the mechanism of expression in vivo as well.

Oroszlan's group has addressed the problem of suppression in MMTV as was done for MLV: protein purification and amino acid sequencing. HIZI et al. (1987) purified from MMTV virions a protein termed p30, suspected to be the C-terminal cleavage product of the *gag-pro* fusion, $p110^{gag\text{-}pro}$. Indeed, upon determining the entire amino acid sequence of p30, these workers could demonstrate that the protein was encoded by both the *gag* and *pro* genes and identify the position on the mRNA at which the reading frame switched. At one of two adjacent codons within the *gag-pro* overlap, the amino acid sequence indicated, ribosomes shifted by one nucleotide in the 5′ direction, moving from the *gag* frame into the *pro* frame. The nature of this frameshift site will be discussed in more detail below.

4 Mechanistic Considerations

The experiments discussed in the previous section have at once solved the *gag-pol* problem and introduced another set of problems altogether. What features of retroviral or retrotransposon mRNA allow or encourage such high-level suppression? The spontaneous rate of frameshifting (at least in *E. coli*) is approximately 3×10^5 per codon (KURLAND 1978), and yet at certain codons in MMTV and TY mRNA frameshifting occurs at the staggering frequency of one in four or five. Termination suppression occurs at the end of the MLV and FeLV *gag* genes at least one hundred times more often then at typical stop codons (CAPONE et al. 1986). What *trans*-acting factors, cellular or viral, are involved in these processes? While our understanding of these issues is far from complete, there has been some progress recently. Where appropriate, I will compare these fledgling models for suppression in retroviral genes with those emerging in other systems, particularly in *E. coli*.

4.1 *cis*-Acting Sequences

Identification of Suppression Sites. By deducing the site of a suppression event, one can potentially learn a great deal about the event itself. This does not apply to termination suppression, of course; by definition suppression occurs at the terminator. In contrast, productive frameshifting can occur at any point along the mRNA where the involved genes overlap. In some cases the overlaps between retroviral genes are greater than 200 nucleotides in length (see Table 1).

The only definitive method of localizing a frameshift site is to sequence the relevant portion of the "trans-frame" protein (defined as a protein encoded by at least two overlapping open reading frames via ribosomal frameshifting). As discussed above, HIZI et al. (1987) used this approach to localize the point of transition from the *gag* to *pro* frames of MMTV to either an AAC-asparagine or a UUG-leucine codon in the *gag* frame within the *gag-pro* overlap. (The prsence of an overlapping leucine codon in the *pro* frame leads to this ambiguity.) Three other -1 frameshift sites have been deduced by amino acid sequencing, all from trans-frame proteins synthesized in vitro in a rabbit reticulocyte system. JACKS and coworkers (1988a,b) have cloned portions of the *gag-pol* overlaps of RSV and HIV-1 downstream of an initiator methionine and a short leader sequence. Translation of mRNA transcribed in vitro from these templates would be expected to produce trans-frame proteins whose N-termini would be within 15 amino acids of the sites of frameshifting. The mRNAs were translated in the presence of several different radioactive amino acids, and the amino acid sequence of the purified proteins deduced from the radioactivity profile of the products of progressive Edman degradation. These analyses identified the same type of codon, a UUA-leucine, as the frameshift site in both the RSV and HIV-1 *gag-pol* overlaps. Amino acid sequencing of the product of a functional point mutant in the RSV frameshift site demonstrated that frameshifting will also occur at a UUU-phenylalanine codon in this context (JACKS et al. 1988b).

Amino acid sequence information is not yet available for the TY transframe protein. However, deletion analysis has implicated a short sequence within the *tya-tyb* overlap. This 11-nucleotide sequence is conserved between the otherwise fairly divergent types of TY elements, TY-1 and TY-2 (WILSON et al. 1986). Furthermore, a 14-nucleotide sequence containing these 11 nucleotides is sufficient to direct + 1 frameshifting when placed in an unrelated mRNA (CLARE et al. 1988). Thus, this sequence appears to be necessary and sufficient for + 1 frameshifting in yeast cells. Precisely where or by what mechanism the frameshift event occurs must await amino acid sequencing and more detailed mutational analysis.

Heptanucleotide Consensus Sequences for − 1 Frameshift Sites. The study of the − 1 frameshift events has benefitted from the numerous documented or suspected examples (see Fig. 2). Even before they were confirmed by amino acid sequencing, the AAC-asparagine codon in the MMTV *gag-pro* overlap and the UUA-leucine codons in the RSV and HIV-1 *gag-pol* overlaps were suspected to be involved in frameshifting, simply because these same codons are found in the overlaps of

several other retroviruses and some retrotransposons. In fact, as shown in Table 1, the overlaps of all of the elements in Classes II and IV of Fig. 2 contain one of the three following sequences: U UUA, U UUU, or A AAC, where the triplet is a codon in the upstream open reading frame. [While the U UUU sequence has not been shown to be the site of frameshifting in any of its native contexts, this sequence will substitute for the natural RSV site (JACKS et al. 1988b; see above).]

The similarity between the overlap sequences actually extends upstream of these putative frameshift sites. In every case save one, these sites are preceded by runs of three A, U, or G residues (Table 1). (The U UUA sequence in the MMTV *pro-pol* overlap, the one exception, is preceded by the sequence GGA.) These similarities suggest that the − 1 frameshift signals encompass seven nucleotides: two adjacent codons and the nucleotide that precedes them.

A Model for − 1 Frameshifting: Simultaneous Slippage. The conserved structure of the documented or suspected − 1 frameshift sites suggests how they might function (JACKS et al. 1988b). This model is shown for the RSV sequence A AAU UUA in Fig. 3. Normal translation delivers the ribosome to the conformation depicted in step I: the AAU codon resident in the P site complexed with $tRNA^{Asn}$ carrying the nascent peptide and the adjacent UUA codon decoded by $tRNA^{Leu}$ in the A site. Slippage by both tRNAs by one nucleotide in the 5′ direction leads to the conformation shown in step II with both tRNAs paired to the overlapping *pol* frame codons, AAA and UUU. Assuming conventional Watson-Crick base-pairs between the tRNA anticodons and their *gag*-frame codons, this *pol*-frame pairing

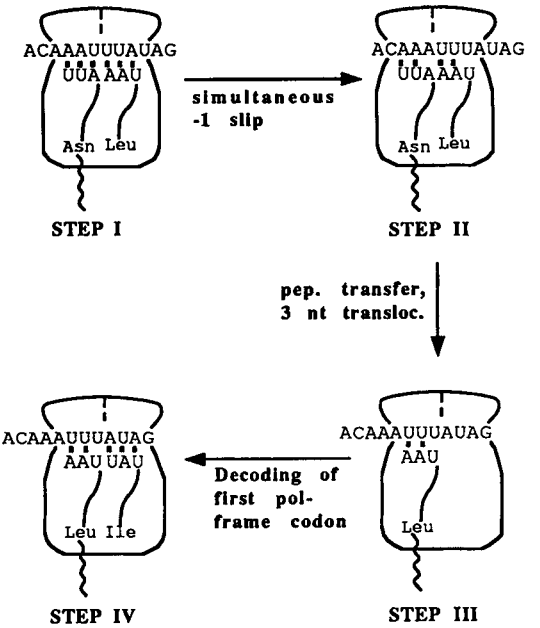


Fig. 3. Simultaneous slippage model for frameshifting. *Step I*: Peptidyl-$tRNA^{Asn}$ and aminoacyl-$tRNA^{Leu}$ are bound to the *gag*-frame codons AAU and UUA. *Step II*: Slippage by both tRNAs by one nucleotide in the 5′ direction results in mispairing to the *pol*-frame codons AAA and UUU. *Step III*: Peptidyl transfer and 3-nucleotide translocation brings the first decoded *pol*-frame codon, AUA, into the ribosomal A site. *Step IV*: Entry of the aminoacyl-$tRNA^{Ile}$ into the A begins translation in *pol*. This model is illustrated for the RSV *gag-pol* frameshift site, but all of the other heptanucleotide sequences shown in Table 1 could substitute for it. Adapted from JACKS et al. (1988b)

would involve only the first two anticodon positions of each tRNA. Peptidyl transfer and three-nucleotide translocation then brings the $tRNA^{Leu}$ (and the nascent peptide) to the P site, delivering the *pol*-frame codon AUA-isoleucine to the A site (step III). Normal translation in the *pol* frame begins with the decoding of the AUA codon by $tRNA^{Ile}$ (step IV). The other heptanucleotide sequences shown in Table 1 could substitute for the RSV sequence in Fig. 3. Although in some cases different tRNAs species would be required, the basic mechanism of single nucleotide slippage by adjacent tRNAs is maintained.

Evidence in Favor of the Simultaneous Slippage Model. Support for the simultaneous slippage model has come not only from the amino acid sequences discussed above but also from mutational analysis of several frameshift sites. The −1 slippage at the so-called A-site codon (for example, the RSV UUA codon; Fig. 3) is consistent with the amino acid sequences of the trans-frame proteins of MMTV, RSV, and HIV-1 (HIZI et al. 1987; JACKS et al. 1988a,b). These sequences implicate the predicted *gag*-frame codons and show that the first *pol*-frame codon decoded is that which directly overlaps the A-site codon. Certain alternative mechanisms for frameshifting at this site would predict different amino acid sequences. For example, if the tRNA reading the A-site codon were to translocate five nucleotides instead of the normal three or slip by two nucleotides in the 3′ direction, the overlapping *pol*-frame codon would be bypassed, and the first decoded *pol*-frame codon would be the next one in line. However, the amino acid sequence analysis alone cannot confirm the -1 slippage model. A mechanism such as 2-nucleotide translocation by the A-site tRNA would also predict the observed amino acid sequences.

The simultaneous −1 slippage model of frameshifting in retroviral overlaps (Fig. 3) is more strongly supported by the effects of point mutations in the frameshift sites in the RSV and HIV-1 *gag-pol* overlaps. For both of these viruses, frameshifting occurs at UUA codons preceded by another U residue. The proposed model predicts that all three of the U residues in the sequence U UUA are necessary for frame shifting as part of the 0- or -1-frame codons bound by the $tRNA^{Leu}$ before and after slippage (Fig. 3). Indeed, mutation of any of the U residues in this sequence in the RSV frameshift site to any other nucleotide eliminates production of the *gag-pol* protein in vitro (JACKS et al. 1988b). Frameshifting in the HIV-1 overlap is also abolished if either of the first two U residues of the U UUA sequence are changed to C or the final U to any nucleotide (JACKS et al. 1988a; WILSON et al. 1988).[2]

The evidence cited above quite convincingly established simple tRNA slippage as the mechanism by which ribosomes are redirected into the -1 frame

[2] Interestingly, mutations of the A position in the U UUA sequences of RSV and HIV-1 are not inhibitory, and, in fact, changing the A to U causes an approximately twofold increase in activity in both cases (JACKS et al. 1988b; WILSON et al. 1988). These results suggest that in addition to $tRNA^{Leu}$, $tRNA^{Phe}$ can also mediate frameshifting along a run of U residues in these contexts. Indeed, amino acid sequencing of a trans-frame protein produced by the RSV A-to-U mutant has shown a phenylalanine residue at the transition from the *gag* to *pol* frames. The sequence U UUU is also thought to be the naturally occurring frameshift site in several retroviral genes (see Table 1)

during frameshifting in retroviral overlaps. The claims that the responsible tRNAs are in the ribosomal A site when the slip occurs and that this slip is coupled with a similar one by the adjacent P-site tRNA (Fig. 3) are less well grounded. As predicted by the model, mutations in the run of three A residues that precede the RSV U UUA sequence do inhibit frameshifting in vitro (by approximately 80%), and a mutation affecting nucleotides directly preceding these A residues has no obvious effect on frameshifting (JACKS et al. 1988b). Furthermore, mutations in the *gag* termination codon, which directly follows the UUA codon, also do not inhibit frameshifting in vitro. These mutations might have been expected to influence frameshifting if the $tRNA^{Leu}$ were to slip into the P site rather than in the A site. Also, mutations in the central position of the HIV-1 P-site codon UUU strongly inhibit frameshifting. Finally, the mere conservation of the heptameric sequence motif in all Class II and Class IV overlaps is very suggestive that two adjacent codons are involved in the process. Nevertheless, disruption of the run of A residues upstream of the RSV U UUA sequence does not abolish frameshifting (these mutants function at approximately 20% the wild-type activity), suggesting that slippage by the P-site tRNA may not be obligatory during the process of frameshifting but might merely facilitate slippage by the A-site tRNA.

Other Examples of tRNA Slippage in Frameshifting. tRNA slippage along homopolymeric sequences has been proposed to account for frameshifting in a number of systems. Stretches of U residues have been suggested as the sites of frameshifting in gene 10 of bacteriophage T7 (DUNN and STUDIER 1983) and in leaky + 1 and − 1 frameshift-mutant alleles of the yeast mitochondrial gene *oxi*1 (FOX and WEISS-BRUMMER 1980). Frameshifting in the release factor II (RFII) gene of *E. coli* has been shown to involve mispairing of the tRNA reading the last 0-frame codon with the overlapping + 1-frame codon (WEISS et al. 1988). WEISS and co-workers (1987) have also demonstrated tRNA slippage by one or more nucleotides in both the 5′ and 3′ directions along numerous synthetic homopolymeric sequences in *E. coli.*

Unlike the simultaneous slippage model for frameshifting in retroviral overlaps, none of these examples is thought to involve slippage by tRNAs at both ribosomal sites. In fact, both for the RFII gene and the synthetic homopolymeric sequences in *E. coli*, the positioning of stop codons adjacent to the frameshift sites greatly increases the frameshifting efficiency (WEISS et al. 1987). This finding suggests that tRNA slippage occurs in the ribosomal P site, while the terminator is in the A site. The positive effect of the stop codons on frameshifting in these settings could result from their extending the time that the frameshift-mediating tRNAs are in the P site (due to slower decoding of the terminator), thereby increasing the probability of P-site tRNA slippage. A different mechanism for achieving this end during frameshifting in retroviral genes is discussed below.

Presumed + 1 Frameshift Sites. In the yeast transposable elements TY-1 and TY-2 the frameshift sites have been grossly defined by the observation that these two elements shared a common 11-nucleotide sequence in their *tya-tyb* overlaps

(WILSON et al. 1986) and, more persuasively, by the demonstration that a 14-nucleotide sequence containing this 11-mer allows frameshifting when placed in a heterologous genetic context (CLARE et al. 1988). The conserved sequence, 5'U CUU AGG CCA C3' (where triplets denote codons in the *tya* frame), is clearly not related to the heptanucleotide frameshift sites described above. Despite the difference in the polarity of the frameshift events, one might have expected that Ty elements possess a similar motif, with the nucleotide sequence arranged to allow tandem tRNAs to shift into the +1 frame. However, this sequence is not indicative of slippage by even one tRNA, nor, in fact, of any alternative mechanism.

Experiments performed in *E. coli* might shed some light on the mechanism of frameshifting with TY RNA. Recently, SPANJAARD and VAN DUIN (1988) observed high-level +1 frameshifting during translation of introduced, adjacent AGG-arginine codons in an otherwise normal mRNA in *E. coli* cells. They postulate that the low abundance of the *E. coli* tRNA isoacceptor that reads AGG results in failure to decode this AGG-AGG doublet properly. Similarly, WEISS and GALLANT (1983) and ATKINS et al. (1979) have reported frameshifting in *E. coli* cells or in vitro translation extracts upon alterations in the concentrations of various charged tRNA species. Frameshifting in these contexts could be a result of improper pairing of a noncognate tRNA in the vacant ribosomal A site (ATKINS et al. 1979) or pairing by a cognate tRNA to an out-of-frame codon (WEISS and GALLANT 1983). In all cases, increasing the concentration of the tRNA corresponding to the "hungry codon" inhibits or eliminates frameshifting. The putative frameshift site in TY may function analogously, since both the CUU and AGG codons (found within the implicated 11-nucleotide sequence) are rarely used in yeast cells and, therefore, probably have correspondingly rare tRNAs (BENNETZEN and HALL 1982).

Although the LINE element L1Md is listed in Fig. 2 as a second example of an element with a directly +1 overlapping *pol* gene, there is as yet no direct evidence that this element utilizes frameshifting to produce a fusion protein. Given the sequence of the 14-nucleotide overlap region in L1Md (LOEB et al. 1986), it is not evident where or how frameshifting would occur there. None of the five codons in the upstream (*gag*-like) frame are conspicuously rare, nor is there a long homopolymeric run of nucleotides. Development of an in vitro assay would greatly facilitate study of frameshifting for this element.

4.2 Additional *cis*-Acting Signals: A Role for RNA Secondary Structure in Suppression

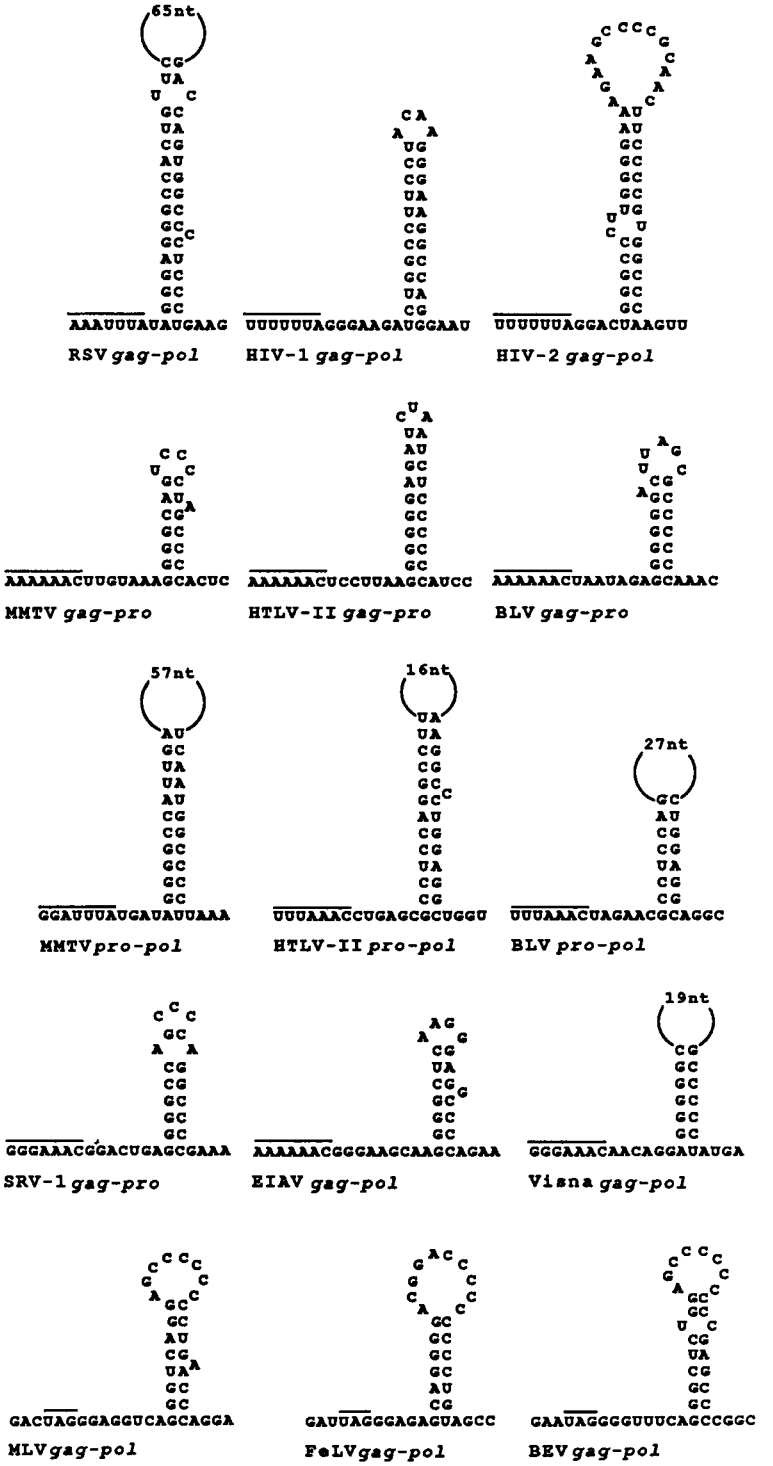
Given that random errors in translation, such as read-through of a termination codon or shift in reading frame, are infequent, the high-level suppression events that occur in retroviral and retrotransposon genes would seem to require specialized signals in the mRNA in order to amplify the frequency of ribosomal

miscues. In part these signals must involve the sequence at the point of suppression, along with its neighboring nucleotide context. I have already discussed the documented and presumed frameshift sites of several retrovirus and retrotransposon genes, and these do encompass sequences several nucleotides in length. While the nucleotide requirements around the *gag* terminators of MLV and FeLV are not known, we presume that the context of the stop codon is equally important here. Indeed, the efficient suppression of UAA and UGA stop codons at the end of MLV *gag* (FENG et al. 1989b) strongly suggests a role for "context" in this case as well. Also, termination suppression in *E. coli* is strongly influenced by the identity of the nucleotide that immediately follows the stop codon (BOSSI 1983; MILLER and ALBERTINI 1983).

There is growing evidence, however, that attention to only the sequence that immediately flank the suppression site might be too narrowly focussed. At least in the case of heptanucleotide -1 frameshift signals (see above), these very sequences appear in the correct reading frame in numerous cellular genes for which there is no evidence (or suspicion) of frameshifting (JACKS et al. 1988b; WILSON et al. 1988). In addition, the heptanucleotide sequences found in the MMTV *gag-pro* and *pro-pol* overlaps (as well as the other nucleotides of the overlaps) are insufficient to direct frameshifting in a novel genetic context (JACKS et al. 1987).

Stem-loop Structures and -1 Frameshifting. The failure of the MMTV *gag-pro* and *pro-pol* overlaps to act in isolation (JACKS et al. 1987) demonstrates that frameshifting efficiency can be affected by sequences outside of the frameshift site. In this case, the negative effect could result from inhibition by surrounding sequences in the nonfunctional mRNAs or the absence of a necessary positive element normally present either upstream or downstream of the frameshift sites in MMTV mRNA. In favor of the latter possibility, JACKS et al. (1987) noted potential stem-loop structures downstream of the two MMTV overlaps. RICE et al. (1985) and SAGATA et al. (1985) had previously called attention to the potential for stem-loop structures downstream of the *gag-pro* overlaps of BLV. In fact, the sequences downstream of all of the putative frameshift sites listed in Table 1 can be folded into stem-loop structures of reasonable stability. A representative set of these is shown in Fig. 4. The structures vary somewhat in the length of the stems and loops, considerably in their base composition, and slightly in the distance between the base of the stem and the frameshift site, but for every retrovirus and retrotransposon known (or suspected) to utilize -1 frameshifting a stem-loop structure can be drawn within 9 nucleotides of the last base of the putative frameshift site.

Direct support for the involvement of RNA secondary structure during ribosomal frameshifting in retroviral genes has come from mutational analysis of the sequences downstream of the RSV frameshift site (JACKS et al. 1988b). Deletion mutations that remove *pol* sequences beginning 23 nucleotides downstream of the base of the stem-loop do not affect frameshifting efficiency in vitro, whereas mutations that remove any or all of the stem structure severely inhibit frameshifting. Interestingly, one mutation that deletes sequences just up to the


65nt
AAAUUUAUAUGAAG
RSV gag-pol
UUUUUUAGGGAAGAUGGAAU
HIV-1 gag-pol
UUUUUUAGGACUAAGUU
HIV-2 gag-pol
AAAAAACUUGUAAAGCACUC
MMTV gag-pro
AAAAAACUCCUUAAGCAUCC
HTLV-II gag-pro
AAAAAACUAAUAGAGCAAAC
BLV gag-pro
57nt
GGAUUUAUGAUAUUAAA
MMTV pro-pol
16nt
UUUAAACCUGAGCGCUGGU
HTLV-II pro-pol
27nt
UUUAAACUAGAACGCAGGC
BLV pro-pol
GGGAAACGGACUGAGCGAAA
SRV-1 gag-pro
AAAAAACGGGAAGCAAGCAGAA
EIAV gag-pol
19nt
GGGAAACAACAGGAUAUGA
Visna gag-pol
GACUAGGGAGGUCAGCAGGA
MLV gag-pol
GAUUAGGGAGAGUAGCC
FeLV gag-pol
GAAUAGGGGUUUCAGCCGGC
BEV gag-pol

base of the RSV stem also inhibits frameshifting, suggesting a possible necessary interaction between sequences in the loop and those downstream of the stem (a so-called pseudo-knot structure; see below). In addition, these deletion mutations were used to demonstrate that a 147-nucleotide sequence from RSV, containing the frameshift site and stem-loop, are sufficient to direct frameshifting in a heterologous genetic context (JACKS et al. 1988b).

To confirm that the RSV stem-loop structure per se is necessary for efficient frameshifting (rather some portion of its primary sequence), JACKS et al. (1988b) constructed site-directed mutations in the stem. Translation of mRNAs containing either of two complementary mutations in the 5′ and 3′ arms of the stem failed to produce any observable Gag-Pol protein. However, when these two mutations were combined in the same mRNA, returning the potential for base-pairing, frameshifting was restored to approximately 50% of the wild-type level.

Given the demonstrated need for mRNA secondary structure during frameshifting in the RSV *gag-pol* region, a similar requirement in other relevant retroviral and retrotransposon genes would seem likely. At least for the −1 frameshift events, the similarity between the putative frameshift sites (Table 1) suggests a conservation of mechanism, and, as discussed above, all of these frameshift sites are followed by sequences that could assume a secondary structure (see Fig. 4). However, experiments performed on HIV-1 show that high-level frameshifting can occur in at least some of these frameshift sites in the absence of obvious downstream secondary structure. MADHANI et al. (1988) constructed a large series of mutations in the region downstream of the HIV-1 frameshift site. For all but one of these mutations, the in vitro frameshifting activity was indistinguishable from wild type. It is not clear why the single mutation had an inhibitory effect. Similarly, WILSON et al. 1988 reported high-level frameshifting in vitro and in yeast cells on a short HIV-1 sequence that does not include the nucleotides involved in the potential stem-loop structure. Thus, at least for HIV-1, a stem-loop structure is not required for efficient frameshifting in vitro or in vivo.

One possible explanation for the differing requirements for different retroviruses has recently arisen from the work of BRIERLEY et al. (1989) concerning a different type of virus altogether. These workers, who had previously provided evidence that the coronavirus avian infectious bronchitis virus (IBV) utilizes frameshifting (BRIERLEY et al. 1988; see also Sect. 6), have now carefully defined the sequences necessary for efficient frameshifting in vitro. In addition to the nucleotides of the presumed frameshift site, UUUAAAC (a site also seen in several retroviral overlaps; Table 1), approximately eighty nucleotides immediately downstream are also required. From the previous work on retroviruses described above, one would assume that these downstream sequences would form a stem-loop structure. However, by constructing numerous mutations and

Fig. 4. Potential stem-loop structures located downstream of retroviral frameshift and termination suppression sites. Predicted stem-loop structures are shown relative to known or suspected sites of suppression (*overlined*). References for nucleotide sequences are found in the legend to Fig. 2

compensatory mutations, BRIERLEY et al. (1989) have shown quite convincingly that the notion of a simple stem-loop is incorrect. Rather, this downstream region must fold into a more complex three-dimensional structure, most likely a pseudo-knot, for efficient frameshifting to proceed.

This finding may help explain two remaining questions about retroviral frameshifting. First, for RSV, where the potential exists for base-pairing between loop nucleotides and sequences downstream of the stem, a requirement for pseudo-knot formation would explain the inhibitory effects of a deletion mutation that leaves the basic stem-loop structure intact (JACKS et al. 1988b; see above). BRIERLEY et al. (1989) have also noted that while many other proposed retroviral stem-loop structures could also form pseudo-knot structures, the proposed HIV-1 structure cannot. Perhaps the HIV-1 frameshift sequence has evolved to the point where a contribution from the downstream structure is not required, and the potential stem-loop structure present there is either unrelated to frameshifting or is a vestigial remnant of a former pseudo-knot. It is also possible that the less energetically stable simple stem-loop structure subtly enhances frameshifting efficiency from what is already a particularly "leaky" frameshift site.

Status of Secondary Structure in +1 Frameshifting and Termination Suppression. Given that retroviruses and retrotransposons do utilize similar strategies to express their *gag*-related gene products, and since they all presumably evolved from some primordial "retro-element", we might expect that the various types of translational suppression would be mechanistically related. However, as already discussed above, this does not seem to hold when comparing the structure of the putative -1 and $+1$ frameshift sites. Also, there is no evidence that mRNA secondary structure is required for frameshifting in the TY-1 overlap (CLARE et al. 1988).

On the other hand, recent experiments addressing the sequence requirements for suppression of the MLV *gag* terminator have suggested that this event may also be dependent on some type of mRNA structure in the neighboring *pol* sequence. FELSENSTEIN and GOFF (1989), assaying MLV termination suppression in vivo, found inhibitory effects by mutations in several nucleotide positions downstream of the *gag* terminator. Although not yet conclusive, these results suggest that the necessary *pol* sequences assume a required secondary (or tertiary) structure. As shown in Fig. 4, potential stem-loop structures exist downstream of the *gag* terminator in the three viruses known or believed to use termination suppression.

Possible Functions for Stem-loop Structures. The presence of an adjacent stem-loop structure could enhance suppression efficiency in several ways. For frameshift suppression, the downstream structure might actually force a fraction of ribosomes at the frameshift site into the -1 frame. The structure (either the stem or loop) might be the binding site for a ribosomal protein, soluble translation factor, or ribosomal RNA. This interaction could destabilize the codon-anticodon interaction which would promote tRNA slippage or mispairing of a tRNA to a termination codon. Arguing against any sequence-specific

interaction, though, is the lack of primary sequence similarity between the various stem-loops (Fig. 4).

A downstream stem-loop structure could also function by simply slowing translation through the suppression site, allowing increased time for the suppression event to occur. In the case of RSV *gag-pol* suppression, the presence of the stem-loop does cause a subset of ribosomes to pause at or near the frameshift site (JACKS et al. 1988b). Translational time-course experiments performed on various RSV mRNA have shown a distinct but transient protein species that comigrates with the expected product of pausing at the frameshift site. The abundance of this "pause product" is greatly reduced when the time-course is performed on an mRNA in which the stem structure has been perturbed. The effect of pausing could be to broaden the time window during which tRNA slippage could occur, thereby increasing the likelihood that a frameshift would have taken place prior to the ensuing tRNA translocation. Pausing at a stem-loop structure could increase the efficiency of termination suppression if the position of the paused ribosome precluded entry by the release factor but not the suppressor tRNA.

The concept of increased "error" with decreased translation rate runs counter to the generally accepted notion that accuracy is sacrificed for increased speed of translation (YARUS and THOMPSON 1983). Several lines of evidence suggest that the need for rapid protein synthesis prevents ribosomes from exercising their full potential to discriminate between cognate and noncognate tRNAs. Reducing the rate of translation by drugs or ribosomal mutations can decrease the frequency of missense errors (THOMPSON and KARIM 1982; THOMPSON 1988). However, YARUS and THOMPSON (1983) have pointed out that errors requiring kinetically slow reactions might be enhanced if translation itself were slowed. Thus, frame maintenance and proper termination may be normally achieved, at least in part, by limiting the time that ribosome-bound tRNAs have to sample the alternative reading frames or for potential suppressor tRNAs to access a termination codon. This hypothesis could be tested directly by examining the effects on translational suppression by agents that artificially slow translation.

4.3 *trans*-Acting Factors

In addition to the *cis*-acting sequences at the suppression sites and possible nearby structures, translational suppression in retroviral and retrotransposon genes must be dependent on certain *trans*-acting factors. At the very least, the tRNA species that carry out the suppression events are necessary conspirators. Specialized factors, viral or cellular, ribosomal proteins, and ribosomal RNAs could also potentially be involved.

Suppressor tRNAs. To date the only implicated *trans*-acting factor for any of these suppression events is a rare glutamine tRNA species able to suppress amber stop codons. This tRNA was isolated by KUCHINO et al. (1987) by virtue of its

ability to suppress efficiently the amber terminator at the end of the coat gene of tobacco mosaic virus. Interestingly, the level of this tRNA species is significantly higher in mouse NIH 3T3 cells that are infected with MLV compared with uninfected 3T3 cells, suggesting that MLV infection might specifically induce expression of the gene for this tRNA. This is an intriguing result since it implies that the virus actively promotes translational suppression rather than simply providing the necessary *cis*-acting sequences and relying on the host for the rest.

The result is also surprising for several reasons. First, it has been known for some time that in vitro translation of MLV vRNA results in a Gag to Gag-Pol ratio that approximates the ratio observed in infected cells (KERR et al. 1976; MURPHY and ARLINGHAUS 1978). FENG et al. (1989a) have recently used a similar in vitro assay to compare the amount of suppressor tRNA activity in normal and MLV-infected cells, and they find no difference between the two cell types. PANGANIBAN (1988) has constructed a vector for assaying termination suppression in vivo containing approximately 300 nucleotides surrounding the *gag* terminator of AKV (a mouse retrovirus derived from an endogenous retrovirus harbored by AKR mouse strains and closely related to MLV). Introduction of this vector into several cell types resulted in approximately 10% suppression of the amber terminator. Significantly, NIH 3T3 cells infected with an amphotropic murine retrovirus did not show an increased level of suppression. Finally, high-level suppression also occurs at both UAA and UGA stop codons placed at the end of MLV *gag* (FENG et al. 1988b). While the glutamine tRNA proposed to be induced by MLV infection might also act on a UAA terminator, it is unlikely to account for UGA suppression (FENG et al. 1989b). Thus, if MLV infection does induce expression of the relevant glutamine tRNA species, this induction appears superfluous, at least in the systems in which it has been studied to date. Perhaps increased production of the suppressor tRNA is only necessary during infection of certain cell types or in the context of the whole animal where host antiviral factors might otherwise limit suppression frequency.

None of the frameshift-mediating tRNAs has been isolated to date. The fact that retroviral Gag-fusion proteins have been detected in several cell types and in vitro translation systems suggests that the tRNAs involved in -1 frameshift are widely distributed. It is interesting that all of the putative -1 frameshift sites include one of three A-site codons: UUU, UUA, or AAC (Table 1). Perhaps the tRNAs that decode these codons are particularly suited for slippage.

Frameshift-suppressor tRNAs have been detected following genetic selection in bacteria and yeast (RIDDLE and ROTH 1970; ROTH and CARBON 1973; ROTH 1974; KOHNO and ROTH 1978; BOSSI and ROTH 1981; GABER and CULBERTSON 1984; BOSSI and SMITH 1984). The most common of these suppressors have an extra nucleotide in the anticodon loop and seem to function by occupying four message nucleotides instead of the normal three, forcing the ribosome into the $+1$ reading frame. This type of RNA could function to suppress the $+1$ frameshift in TY-1, although they have not been observed in wild-type yeast strains. One -1 frameshift suppressor tRNA has been characterized from *Salmonella* (D. J. O'MAHONEY et al. 1988, unpublished observations). It lacks one

of the normal anticodon loop nucleotides and is thought to cause −1 frameshifting by translocating, or otherwise occupying, just two message nucleotides. This suppressor is probably not a good model for tRNAs that cause −1 frameshifting by slippage, however.

Non-tRNA Factors. In order to understand suppression mechanisms at the molecular level, one must first identify all of the players. In addition to the relevant codons and tRNAs that read them and the other *cis*-acting mRNA sequences, suppression certainly involves other factors. Ribosomal proteins, for example, are known to affect the fidelity of translation (STRINGINI and GORINI 1970; ROSSET and GORINI 1969). Mutant elongation factors may also enhance the frequency of termination suppression (CULBERTSON et al. 1982). Frameshifting in the RFII gene of *E. coli* requires an interaction between 16*S* ribosomal RNA (the sequence that normally recognizes the Shine-Dalgarno sequence during translational initiation) and the mRNA sequence just upstream of the frameshift site (WEISS et al. 1988). Defining the additional factors involved in suppression in retroviral and retrotransposon genes may require establishing a genetic selection in which, for example, cell viability is dependent on a translational suppression event. Such a selection might best be carried out in a genetically tractable system like bacteria or yeast, assuming that the suppression event of interest occurs in that system.

With the possible exception of virus infection raising the level of suppressor tRNA (KUCHINO et al. 1987; see above), viral proteins do not seem to be required for termination suppression or frameshifting. The efficiencies of suppression observed on mRNAs in vitro (including mRNAs from which no viral products could be produced) appear to rival the in vivo levels. Also, termination suppression on a short sequence from AKV occurs equally well in virus-infected and uninfected cells (PANGANIBAN 1988).

5 Physiological Effects

In this section I have included those subjects that relate more to the consequences of translational suppression rather than its mechanism.

Efficiency of Suppression. Through the analysis of various normal and mutant suppression sites, it has become clear that a wide range of suppression efficiencies are possible. For example, frameshifting occurs in the MMTV *gag-pro* overlap at approximately 25% efficiency (JACKS et al. 1987; MOORE et al. 1987), while a point mutation in the RSV frameshift site reduces the efficiency there to about 1% (JACKS et al. 1988b). Thus, depending on the exact nature of the *cis*-acting sequences at the suppression site, the relative amount of the product of suppression could be anywhere from 1 part in 4 to 1 in 100. Why then did the different sequences (and their corresponding efficiencies) evolve in the different

viruses?[3] In some instances, one can make a reasonable guess. For those viruses that require two successive frameshift events to access their *pol* genes [the Class IV viruses with *pro* genes intervening between *gag* and *pol* (Fig. 2)], it is expected that at least the first of them should be quite efficient or else very little of the Gag-Pro-Pol protein would be produced. But why, for example, the frameshifting efficiencies in the RSV and HIV-1 *gag-pol* overlaps should be 5% and 10%, respectively, and not 25% is not known.

The issue of suppression efficiency is particularly interesting because the available evidence suggests that the ratio of Gag to its fusion proteins may strongly influence virus replication. FELSENSTEIN and GOFF (1988) have shown that a nonfunctional mutant of MLV in which the *gag* terminator has been converted to a glutamine codon is only weakly rescued by the expression of an exogenous *gag* gene. The implication of this result is that the normal 20:1 Gag to Gag-Pol ratio is necessary for maximal MLV virion production. For RSV, P. PRYCIAK et al. (1988, unpublished observations) found inhibition of virus production by mutations previously shown to affect in vitro frameshifting efficiency. RSV production is impaired not only by mutations that eliminate frameshifting in vitro, but also by only partially inhibitory and one partially stimulatory mutation. Thus, even subtle alterations in the ratio of Gag to Gag-Pol can have significant effects. Perhaps due to the geometry of the viral capsid, a proper ratio of Gag to its fusion proteins is necessary during virus assembly. The different suppression efficiencies observed for different viruses may reflect subtle differences in the ways in which their core subunits are assembled.

Affecting Suppression Efficiency as a Means of Virus Inhibition. If virus replication is sensitive to subtle changes in the ratio of Gag to its fusion proteins, it might be possible to block virus production with agents that either inhibit or stimulate suppression frequency. KUCHINO et al. (1988) have recently reported that Avarol, a substance isolated from the sponge *Dysidea avara*, inhibits the MLV-infection-induced expression of the glutamine tRNA species thought to suppress the *gag* terminator. Thus, the observed inhibition of virus production by Avarol may be mediated, albeit indirectly, by an inhibition of termination suppression. Other, more direct inhibitors or stimulators of suppression frequency could be imagined, but none has as yet been described. Such an agent would be potentially valuable as an inhibitor of HIV-1, especially if the cellular side effects were limited. It is not currently known whether any eukaryotic cellular

[3] A separate, but equally interesting question is how do these different *cis*-acting sequences deliver different frameshifting efficiencies? With the exception of certain cases involving frameshift site mutations (where mutations may disrupt potential base pairing between a tRNA and an alternate-frame codon), the answer is unknown. Efficiency probably results from a combination of the nature of the site structure, and relative abundance of the suppressor tRNAs, and, where applicable, the stability and positioning of the stem-loop structure. Evidence for the first point comes from comparing the frameshifting efficiency of the wild-type RSV *gag* gene with that of a mutant that substitutes the MMTV *gag-pro* frameshift site for the natural one (JACKS et al. 1988b). While this mutant site functions more efficiently than the wild-type one (10% versus 5%), the efficiency is not as high as is obtained on this sequence in its native setting

genes require frameshifting or termination suppression for their expression (see below). However, even if such genes do exist and their expression is necessary for cellular viability, conditions probably exists that affect suppression levels sufficiently to disrupt virus production without causing cellular toxicity.

Significance of the Sites of Suppression. For those elements that utilize termination suppression for *gag-pol* expression or whose frameshift sites correspond to the last codons of the 0-frame (see Table 1), the resulting fusion proteins carry the complete protein sequence encoded by the upstream gene. For example, the Gag moieties of the Gag-Pol proteins of MLV and RSV (whose frameshift site covers the last two *gag* codons) exactly match Gag proteins themselves. However, several putative frameshift sites are not positioned at the end of the upstream genes (Table 1). In these cases, the trans-frame proteins have substituted sequences encoded in the -1 frame of the overlap for the sequences normally present at the C-terminus of the uni-frame protein. Again, by way of example, the HIV-1 frameshift site is located very near the 5′-end of the 205-nucleotide *gag-pol* overlap (JACKS et al. 1988a). As such, the final 65 aa of the Gag protein are not present in the Gag-Pol fusion; rather the amino acids encoded by the last 65 *pol*-frame codons of the overlap are in their place. Whether this amino acid sequence difference is functionally significant, though, is unclear. Indeed, LOEB et al. (1989) have argued that most of the *gag-pol* sequence encoded by the *pol*-frame codons of the overlap are functionally unimportant, since there is little sequence similarity in this region between the two isolates of HIV, HIV-1 and HIV-2. These workers have suggested that the *pol*-encoded sequences may merely serve as a spacer between the Gag and protease domains in the Gag-Pol protein. (The N-terminus of the protease is encoded near the 3′-end of the *gag-pol* overlap.) Another intriguing suggestion is that these sequences (and, by extension, others similarly located) may be maintained because they serve a necessary function in the suppression event itself, such as forming a portion of a stem-loop structure (LOEB et al. 1989; see Fig. 2 and Sect. 4.2).

6 Additional Examples, Counter-Examples, and Future Examples

In the introduction to this review, I discussed the history of the *gag-pol* problem, particularly the emergence of an consensus solution that was based in large part on preconceptions about how eukaryotes controlled their genes. The recent discovery of translational suppression as the actual solution to the *gag-pol* problem in, at least, several cases has increased our appreciation of the variety of available genetic control mechanisms. This broadened view has and will continue to aid in the discovery of additional examples of translational suppression in the control of eukaryotic gene expression. Ironically, though, the knowledge that some retroviruses and retrotransposons utilize translational suppression mech-

anisms might also have fostered the belief that all *gag-pol*-like genes will be controlled in a similar fashion. This preconception, radically different from the one that slowed progress in the understanding of the *gag-pol* problem only a few years ago, is probably incorrect as well.

Counter-examples. There are classes of "retro-elements" that seem to express their *pol* genes without the use of termination suppression or ribosomal frameshifting. For the retrotransposons *copia* of *Drosophila*, *Tal* of *Arabidopsis*, and *Tnti* of tobacco, synthesis of a Gag-Pol protein requires only standard translation (MOUNT and RUBIN 1985, VOYTAS and AUSUBEL 1988; GRANDBASTIEN et al. 1989). As shown in Fig. 2, these transposons carry both the *gag* and *pol* coding domains in one continuous open reading frame and, thus, present an interesting twist on the problem: How to express the Gag protein alone. Possibilities include ribosomal frameshifting near the end of the *gag* domain or cleavage of some fraction of the Gag-Pol protein prior to core assembly. It is also possible that the Gag-Pol protein is sufficient for core assembly. However, the most likely explanation is that these elements produce a separate, *gag*-specific message. An mRNA species seemingly containing only the *gag* portion of the *copia gag-pol* gene has been observed in *Drosophila* cells (FLAVELL et al. 1981). Such an mRNA could be synthesized either by premature transcriptional termination or mRNA splicing.

HBV and CaMV, generally considered DNA viruses, require reverse transcription in their life cycles. The *pol* genes of these viruses all lie downstream of and overlap (in the + 1 direction) the genes encoding the viral core proteins. Due to the conservation of this overlapping structure and the use of ribosomal frameshifting by the related retro-elements, one might expect that these DNA viruses would express a core-Pol fusion protein via + 1 ribosomal frameshifting. However, in vitro translation of mRNAs containing the CaMV (GORDON et al. 1988) and HBV (CHANG et al. 1989) core-*pol* overlaps fail to produce the relevant fusion proteins. In addition, recent genetic evidence strongly suggests that CaMV (PENSWICK et al. 1988) and, at least, duck HBV (SCHLICHT et al. 1989; CHANG et al. 1989) express their *pol* genes by internal translational initiation within the *pol* gene, producing a separate Pol protein. Perhaps the differences between the replication strategies of these viruses and RNA viruses obviates the need for a core-Pol protein. Why the core and *pol* genes should overlap, then, is unclear.

Additional Examples of Translational Suppression. Control of expression by translational suppression in eukaryotic cells is not limited to the genes of retroviruses and retrotransposons. Indeed, termination suppression was described by PELHAM (1978) for the coat gene of TMV in 1978. Rattle snake mosaic virus (PELHAM 1979) and two alphaviruses, sindbis virus (STRAUSS et al. 1983) and Semliki forest virus (STRAUSS et al. 1984), are also believed to utilize termination suppression.

Ribosomal frameshifting has been proposed by BRIERLY et al. (1987) to account for expression of a long open reading frame (F2) of the avian coronavirus infectious bronchitis virus (IBV). F2 partially overlaps the upstream open reading

frame, F1, in the −1 direction. Although an F1-F2 fusion protein has not been observed in IBV-infected cells, in vitro translation of mRNAs containing the F1-F2 overlap yields proteins consistent with ribosomal frameshifting. In addition, the IBV overlap includes the sequence U UUA AAC [believed to be the frameshift site for several retroviruses (Table 1)], and this sequence is followed closely by a potential stem-loop structure (BRIERLY et al. 1987; see above). Recent mutational analysis has confirmed the requirement for this heptanucleotide sequence and a complex downstream structure in frameshifting in vitro (BRIERLY et al. 1989; see above). Thus, frameshifting in the IBV F1-F2 overlap is almost certainly mechanistically related to frameshifting in retroviral genes.

A computer-assisted search of nucleotide sequence data bases for the putative heptanucleotide frameshift sites listed in Table 1 has uncovered another group of viral genes that may utilize frameshifting (JACKS et al. 1988b). These include genes of tobacco etch virus (ALLISON et al. 1986) and three alphaviruses: sindbis virus (RICE and STRAUSS 1981), semliki forest virus (GAROFF et al. 1980), and Ross river virus (DALGARNO et al. 1983). As yet, there is no independent evidence that frameshifting occurs in any of these genes, however.

Translational Suppression in Cellular Genes. Generally, viral mechanisms, including mechanisms of gene expression, mimic those of the host cell. Thus, with the discovery of translational suppression in certain retroviral and retrotransposon genes came the expectation that cellular examples would quickly follow. At present, however, there is but one example of a frameshift-controlled cellular gene [the RFII gene of *E. coli* (CRAIGEN et al. 1985; see above)] and no known examples of cellular genes that require termination suppression. One explanation for the dearth of cellular counterparts is the relatively small fraction of cellular genes that have been examined at the DNA sequence level. Also, the fact that termination suppression and ribosomal frameshifting have only recently achieved recognition as viable control mechanisms might have led to potential examples being previously overlooked. Two groups have used computer-assisted nucleotide sequence data base searches to find possible eukaryotic frameshift-controlled cellular genes (JACKS et al. 1988b; WILSON et al. 1988). Although several genes were found to contain putative heptanucleotide frameshift sites (see Table 1), none had additional features (downstream secondary structures, extended alternative open readings frames) that would indicate that frameshifting might actually occur there.

One could argue that without the constraints of maintaining a small genome size and with available mechanisms such as alternative mRNA splicing and mRNA editing, the eukaryotic cell has outmoded or never evolved genes that would require this type of translational control. Until the first eukaryotic cellular example is discovered, this position is impossible to refute. Armed with the information gleaned from retroviruses and retrotransposons, however, if such cellular example exist, their discovery should not be far off.

Acknowledgements. I thank Judy Levin, Alan Rein, and Kevin Felsenstein for communicating results prior to publication.

References

Allison R, Johnston R, Dougherty WG (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

Atkins JF, Gesteland RF, Ried BR, Anderson CW (1979) Normal tRNAs promote ribosomal frameshifting. Cell 18: 1119–1131

Bennetzen JL, Hall BD (1982) Codon selection in yeast. J Biol Chem 257: 3026–3031

Bossi L (1983) Context effects: Translation of AUG codon by suppressor tRNA is affected by the sequence following UAG in the message. J Mol Biol 164: 73–87

Bossi L, Roth JR (1981) Four-base codons ACCA, ACCU and ACCC are recognized by frameshift suppressor sufJ. Cell 25: 489–496

Bossi L, Smith (1984) Suppressor sufJ: a novel type of tRNA mutant that induces translational frameshifting. Proc Natl Acad Sci USA 81: 6105–6109

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

Brierly I, Digard P, Inglis SC (1989) Characterization of an efficient ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57: 537–547

Capone JP, Sedivy JM, Sharp PA, Rajbhandary VC (1986) Introduction of UAG, UAA, and UGA nonsense mutations at a specific site in Escherichia coli chloramphenicol acetyltransferase gene: use in measurement of amber, ochre, and opal suppression in mammalian cells. Mol Cell Biol 6: 3059–3067

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

Chang L-J, Pryciak P, Ganem D, Varmus HE (1989) Biosynthesis of the reverse transcriptase of hepatitis B virus involves de novo translational initiation not ribosomal frameshifting. Nature 337: 364–368

Clare J, 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

Clare JJ, Belcourt M, Farabough PJ (1988) Efficient translational frameshifting occurs within a conserved sequence of the overlap between the two genes of a yeast Ty1 transposon. Proc Natl Acad Sci USA 85: 6816–6820

Craigen WJ, Cook RG, Tate WP, Caskey CT (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

Culbertson MR, Gaber RF, Cummins CM (1982) Frameshift suppression in Saccharomyces cerevisiae V. Isolation and genetic characterization of nongroup-specific suppressors. Genetics 102: 361–378

Cummins CM, Culbertson MR, Knapp G (1985) Frameshift suppressor mutations outside the anti-codon in yeast proline rRNAs containing an intervening sequence. Mol Cell Biol 5: 1760–1771

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

Dickson C, Atterwill M (1979) Composition, arrangement and cleavage of the mouse mammary tumor virus polyprotein precursor Pr77*gag* and p110*gag*. Cell 17: 1003–1012

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

Felsenstein KM, Goff SP (1988) Expression of the gag-pol fusion protein of Moloney murine leukemia virus without gag protein does not induce virion formation or proteolytic processing. J Virol 62: 2179–2182

Felsenstein KM, Goff SP (1989) (manuscript in preparation)

Feng Y-X, Hatfield DL, Rein A, Levin JG (1989a) Translational read through of the murine leukemia virus gag gene amber codon does not require virus-induced alteration of tRNA. J Virol 63: 2405–2410

Feng Y-X, Levin JG, Hatfield DL, Schaefer TS, Gorelick RJ, Rein A (1989b) Suppression of UAA and UGA termination codons in mutant murine leukemia viruses. J Virol 63: 2870–2873

Flavell AJ, Levis R, Simon MA, Rubin GM (1981) The 5′ termini of RNAs encoded by the transposable element copia. Nucleic Acids Res 9: 6279–6291

Fox TD, 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 AM, Simons K, Leharch H, Delius H (1980) Nucleotide sequence of cDNA coding for Semliki Forest virus membrane glycoproteins. Nature 288: 236–241

Gordon K, Pfeiffer P, Futterer J, Hohn T (1988) In vitro expression of cauliflower mosaic virus genes. EMBO J 7: 309–317

Grandbastien M–A, Spielmann A, Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337: 376–380

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

Hayward WS (1977) Size and genetic content of viral RNAs in avian oncovirus-infected cells. J Virol 24: 47–63

Hiramatsu K, Nishida J, Naito A, 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 LE, Copeland TD, Sowden RC, Hixson CV, 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, Varmus HE (1985) Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science 230: 1237–1242

Jacks T, Townsley K, Varmus HE, 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

Jacks T, Power MD, Masiarz FR, Luciw PA, Barr PJ, Varmus HE (1988a) Characterization of ribosomal mutations in HIV-1 gag-pol expression. Nature 331: 280–283

Jacks T, Madhani HD, Masiarz FR, Varmus HE (1988b) Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell 55: 447–458

Jamjoon GA, Naso RB, Arlinghaus RB (1977) Further characterization of intracellular precursor polypeptides of Rauscher leukemia virus. Virology 78: 11–34

Kerr IM, Olshevsky U, Lodish HF, Baltimore D (1976) Translation of murine leukemia virus RNA in a cell-free system from animal cells. J Virol 18: 627–635

Kohno T and Roth JR (1978) A Salmonella frameshift suppressor that acts at runs of A residues in the messenger RNA. J Mol Biol 126: 37–52

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

Kuchino Y, Beker H, Akita N, Nishimura S (1987) Natural UAG suppressor tRNA is elevated in mouse cell infected with Moloney murine leukemia virus. Proc Natl Acad Sci USA 84: 2668–2672

Kuchini Y, Nishimura S, Schroder HC, Rottmann M, Muller WEG (1988) Selective inhibition of formation of suppressor glutamine tRNA in Moloney murine leukemia virus-infected NIH3T3 cells by Avarol. Virology 165: 518–526

Kurland CG (1978) Reading frame errors on ribosomes. In: Celis JE, Smith JD (eds) Nonsense mutations and tRNA suppressors. Academic, London pp 98–108

Levin JG, Hu SC, Rein A, Messer LI, Gerwin B (1984) Murine leukemia virus mutant with a frameshift in the reverse transcriptase coding region: implications for pol gene structure. J Virol 51: 470–478

Loeb DD, Padgett RW, Hardies SC, Shehee WR, Comer MB, Edgell MH, Hutchison CA (1986) The sequence of a large L1Md element reveals a tandemly repeated end and several features found in retrotransposons. Mol Cell Biol 6: 168–182

Loeb DD, Hutchison CA, Edgell MH, Farmerie WG, Swanstrom R (1989) Mutational analysis of the HIV-1 protease suggests functional homology with aspartic proteinases. J Virol 63: 111–121

Madhani HD, Jacks T, Varmus HE (1988) Signals for the expression of the HIV pol gene by ribosomal frameshifting. In: Franza R, Cullen B, Wong-Stall F (eds) Control of HIV Gene Expression. Cold Spring Harbor Laboratory, Cold Spring Harbor pp 119–125

Marlor RL, Parkhurst SM, Corces VG (1986) The Drosophila melanogaster gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol Cell Biol 6: 1129–1134

Meitz JA, Grossman Z, Lueders KK, Kuff EL (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 SM, Dobson MJ, Wilson W, Kingsman SM, Kingsman AJ (1985) A retrovirus-like strategy for expression of a fusion protein encoded by the yeast transposon Tyl. Nature 313: 243–246

Miller JH, Albertini AM (1983) Effects of surrounding sequence on the suppression of nonsense codons. J Mol Biol 164: 59–71

Moore R, Dixon M, Smith R, Peters G, 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

Mount SM, Rubin GM (1985) Complete nucleotide sequence of the Drosophila transposable element copia: homology between copia and retroviral proteins. Mol Cell Biol 5: 1630–1638

Murphy EC, Arlinghaus RB (1978) Cell-free synthesis of Rauscher murine leukemia virus "gag" and "gag-pol" precursor polypeptides from virion 35S RNA in a mRNA-dependent translation system derived from mouse tissue culture cells. Virology 86: 329–343

Murphy EC, Campos D, Arlinghaus RB (1979) Cell-free synthesis of Raucher murine leukemia virus "gag" and "env" gene products from separate cellular mRNA species. Virology 93: 293–302

Nam SH, Hatanaka M (1986) Identification of a protease gene of human T-cell leukemia virus type I (HTLV-I) and its structural comparison. Biochem Biophys Res Commun 139:129

Oppermann L, Bishop JM, Varmus HE, Levintow L (1977) A joint product of the genes gag and pol of avian sarcoma virus: a possible precursor of reverse transcriptase. Cell 12: 993–1005

Panganiban AT (1988) Retroviral gag gene amber codon suppression is caused by an intrinsic cis-acting component of the viral mRNA. J Virol 62: 3574–3580

Pawson TP, Martin GS, Smith AE (1976) Cell-free translation of virion RNA from non-defective and transformation-defective Rous Sarcoma viruses. J Virol 19: 950–967

Pelham HRB (1978) Leaky UAG termination codon in tobacco mosaic virus RNA. Nature 272: 469–471

Pelham HRB (1979) Translation of tobacco rattle snake virus RNAs in vitro: four proteins from three RNAs. Virology 97: 256–265

Penswick T, Hubler R, Hohn T (1988) A viable mutation in cauliflower mosaic virus, a retroviruslike plant virus, separates its capsid protein and polymerase genes. J Virol 62: 1460–1463

Philipson L, Andersson P, Olshevsky U, Weinberg R, Blatimore D, Gesteland R (1978) Translation of MuLV and MSV RNAs in nuclease-treated reticulocye extracts: Enhancement of the gag-pol polypeptide with yeast suppressor tRNA. Cell 13: 189–199

Power MD, Marx PA, Bryant ML, Gardner MD, Barr PJ, Luciw PA (1986) Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science 231: 1567–1572

Purchio AF, Erikson E, Erikson RL (1977) Translation of 35S and of subgenomic regions of avian sarcoma virus RNA. Proc Natl Acad Sci USA 74: 4661–4665

Ratner L, Hazeltine W, Patarca R, Livak KJ, Starcich B, Josephs SF, Doran ER, Rafalski JA, Whitehorn EA, Baumeister K, Ivanoff L, Petteway SR, Pearson ML, Lautenberger JA, Papas TS, Ghrayeb J, Chang NT, Gallo RC, Wong-Staal F (1985a) Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 313: 277–284

Ratner L, Josephs SF, Starcich B, Hahn B, Shaw GM, Gallo RC, Wong-Staal F (1985b) Nucleotide sequence analysis of a varient human T-cell leukemia virus (HTLV-Ib) provirus with a deletion in Px-1. J Virol 54: 781–790

Rettenmier W, Karess RE, Anderson SM, Hanafusa H (1979) Tryptic peptide analysis of avian oncovirus gag and pol gene products. J Virol 32: 102–113

Rice CM, Strauss JH (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: 1062–2066

Rice NR, Stephens RM, Burny A, Gilden RV (1985) The gag and pol genes of bovine leukemia virus: nucleotide sequence and analysis. Virology 142: 357–377

Riddle D, Carbon J (1973) A nucleotide addition in the anticodon of a glycine tRNA. Nature 242: 230–234

Riddle DL, Roth JR (1970) Suppressors of frameshift mutations in Salmonella typhimurium. J Mol Biol 54: 131–144

Rosset R, Gorini L (1969) A ribosomal ambiguity mutation. J Mol Biol 39: 95–112

Roth JR (1974) Frameshift mutations. Annu. Rev Genet 8: 319–346

Sagata N, Yasunaga T, Tsuzuku-Kawamura J, Ohishi K, Ogawa Y, 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, Matsu Y, Inouye S, Yoshioka K, 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 MD, Barr PJ, Steimer KS, Stempier MM, Brown-Shimer SL, Gee WW, Renard A, Randolf A, Levy JA, Dina D, Luciw PA (1985) Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2) Science 227: 484–492

Schlicht H–J, Radziwill G, Schaller H (1989) Synthesis and encapsidation of duck hepatitis B virus reverse transcriptase do not require formation of core polymerase fusion proteins. Cell 56: 85–92

Schwartz DE, Tizard R, Gilbert W (1983) Nucleotide sequence of Rous sarcoma virus. Cell 32: 853–869

Seiki M, Hattori Y, Hirayama Y, Yoshida M (1983) Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc Natl Acad Sci USA 80: 3618–3622

Shimotohno K, Takahashi Y, Shimizu N, Gojobori T, Golde DW, Chen ISY, Miwa M, 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 TM, Lerner RA, Sutcliffe JG (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, Wain-Hobson S (1985) Nucleotide sequence of the visna lentivirus: relationship with the AIDS virus. Cell 42: 369–382

Sonigo P, Barker C, Hunter E, Wain-Hobson S (1986) Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell 45: 375–385

Spanjaard RA, van Duin J (1988) Translation of the sequence AGG-AGG yields 50% ribosomal frameshifting. Proc Natl Acad Sci USA 85: 7967–7971

Stacey DW, Allfrey VG, Hanafusa H (1977) Microinjection analysis of envelope-glycoprotein messenger activities of avian leukosis virus RNAs. Proc Natl Acad SCi USA 74: 1614–1618

Stephens RM, Casey JW, Rice NR (1986) Equine infectious anemia virus gag and pol genes: relatedness to visna and AIDS virus. Science 231: 589–594

Strauss EG, Rice CM, Strauss JH (1983) Sequence coding for the alphavirus nonstructural protein is interrupted by an opal terminator codon. Proc Natl Acad Sci USA 80: 5271–5275

Strauss EG, Rice CM, Strauss JH (1984) Complete nucleotide sequence of the genomic RNA of sindbis virus. Virology 133: 92–110

Stringini P, Gorini L (1970) Ribosomal mutations affecting efficiency of amber suppression. J Mol Biol 47: 517–530

Tamura TA (1983) Provirus of M7 baboon endogenous virus, nucleotide sequence of the gag-pol region. J Virol 47: 137–145

Thompson RC (1988) Ef-Tu provides an internal kinetic standard for translational accuracy. Trends Biochem Sci 13: 91–93

Thompson RC, Karim AM (1982) The accuracy of protein biosynthesis is limited by its speed: high fidelity selection by ribosomes of aminoacyl-tRNA ternary complexes containing GTP[γS]. Proc Natl Acad Sci USA 79: 4922–4926

Vaytas DF, Ausubel FM (1988) A copia-like transposable element family in Arabidopsis thalina. Nature 336: 242–244

Von der Helm K, Duesberg PH (1975) Translation of Rous sarcoma virus RNAs in cell-free systems from ascites Kerbs II cells. Proc Natl Acad Sci USA 72: 614–618

Wain-Hobsen S, Sonigo P, Danos O, Cole S, Alizon M (1985) Nucleotide sequence of the AIDS virus, LAV. Cell 40: 9–17

Weiss R, Gallant J (1983) Mechanism of ribosome frameshift during translation of the genetic code. Nature 302: 389–393

Weiss R, Teich N, Varmus H, Coffin J (eds) (1984) RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York

Weiss RB, Dunn DM, Atkins JF, Gesteland RF (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 RB, Dunn DM, Dahlberg AE, Atkins JF, Gesteland RF (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

Weiss SR, Varmus HE, Bishop JM (1977) The size and genetic composition of virus-specific RNAs in the cytoplasm of cells producing avian sarcoma-leukemia viruses. Cell 12: 983–992

Weiss SR, Hackett PB, Oppermann H, Ullrich A, Levintow L, Bishop JM (1978) Cell-free translation of avian sarcoma virus RNA: suppression of the gag termination codon does not augment synthesis of the joint gag/pol product. Cell 15: 607–614

Wilson W, Braddock M, Adams SE, Rathjen PD, Kingsman SM, Kingsman AJ (1988) HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell 55: 1159–1169

Wilson W, Malim MH, Mellor J, Kingsman AJ, Kingsman SM (1986) Expression strategies of the yeast retrotransposon Ty: a short sequence directs ribosomal frameshifting. Nucleic Acids Res 14: 7001–7016

Yarus M, Thompson RC (1983) Precision of protein biosynthesis. In: Beckwith JR, Davies JE, Gallant JA (eds) Gene function in prokaryotes. Cold Spring Harbor Laboratory, Cold Spring Harbor

Yoshinaka Y, Katoh I, Copeland TD, Orozslan SJ (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 TD, Oroszlan SJ (1985b) Translational readthrough of an amber termination codon during synthesis of feline leukemia virus protease. J Virol 55: 870–873