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

ADVANCES IN VIRUS RESEARCH, VOL. 41

TRANSLATIONAL SUPPRESSION IN RETROVIRAL GENE EXPRESSION

Dolph L. Hatfield*, Judith G. Levin†, Alan Rein,‡ and Stephen Oroszlan‡

*Laboratory of Experimental Carcinogenesis
National Cancer Institute
Bethesda, Maryland 20892

†Laboratory of Molecular Genetics
National Institute of Child Health and Human Development
National Institutes of Health
Bethesda, Maryland 20892

‡Laboratory of Molecular Virology and Carcinogenesis
ABL-Basic Research Program
NCI-Frederick Cancer Research and Development Center
Frederick, Maryland 21702

I. Introduction
II. Readthrough Suppression
  A. Role of Context in Readthrough Suppression
  B. Alternate Stop Codons
  C. Suppressor tRNAs
  D. Readthrough Suppression in Other Viruses
III. Ribosomal Frameshifting
  A. Frameshifting in the −1 Direction
  B. Frameshifting in the +1 Direction
IV. Concluding Remarks
  References

I. Introduction

Retroviruses are a unique class of viruses that have been found in all classes of vertebrates, but not in other organisms. Perhaps their most distinctive properties are the flow of information from RNA to DNA early in the infectious process, and the subsequent integration of the viral DNA into the chromosomal DNA of the host cell. Retroviruses are the causative agents of acquired immunodeficiency syndrome (AIDS)

and of a variety of neoplastic diseases in man and domestic animals (Coffin, 1990; Varmus and Brown, 1989).

It is interesting to note that elements with striking similarities to retroviruses, termed retrotransposons, occur in yeast and many other eukaryotes; elements sharing some characteristics with retroviruses have also recently been observed in prokaryotes (reviewed in Garfinkel, 1991). Because of the apparent relationship between retroviruses and retrotransposons, we will consider retrotransposons as well as retroviruses in this review.

In addition to structural proteins, all retroviruses encode at least three enzymes: a protease (PR), which processes the internal proteins of the virion during virus maturation; reverse transcriptase (RT), which copies the genomic RNA of the virion into DNA when the particle infects a cell; and integrase (IN), which catalyzes the insertion of the viral DNA into the chromosomal DNA of the host cell (Dickson *et al.*, 1984). Because these enzymes all function in the free virus particle or in the early stages of infection, they must be incorporated into the virion during virus assembly. However, because they act catalytically, they are needed in much lower amounts than the structural proteins of the virion.

Retroviruses have evolved a remarkable mechanism for expression of the genomic sequences encoding these enzymes. This mechanism, translational suppression, appears to fulfill simultaneously both of the requirements noted above, because it results in a relatively low level of expression of the enzymes and provides a way for the enzymes to be incorporated into the nascent virus particle.

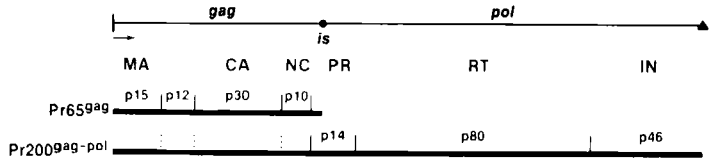
It is now clear that the genomic RNA of the virus is the mRNA for the internal structural proteins of the virus (termed the Gag proteins) and the viral enzymes. As indicated in Fig. 1, the *gag* coding sequences are found at the 5′ end of this mRNA. The enzyme-coding region (generally referred to as the *pol* gene) is immediately 3′ of the termination codon at

---

FIG. 1. Expression of retroviral *pol* and *pro* genes from a single *gag–pro–pol* translational unit by in-frame readthrough, single frameshift, and double frameshift, as illustrated for the respective viruses. The *gag*, *pro*, and *pol* open reading frames are shown together with the following symbols for termination codons: ● UAG; ▲, UAA, and ▼, UGA. The boldface horizontal bars represent the primary translational products, which are processed into smaller functional units as indicated. The protein nomenclature used is that of Leis *et al.* (1988); MA, matrix; CA, capsid; NC, nucleocapsid; PR, protease; RT, reverse transcriptase; and IN, integrase. is, In-frame suppression; and fs, frameshift. Locations of transframe proteins (TF) are also shown. The numbers indicate the approximate molecular weights of the proteins. Arrows indicate the site of translation initiation.

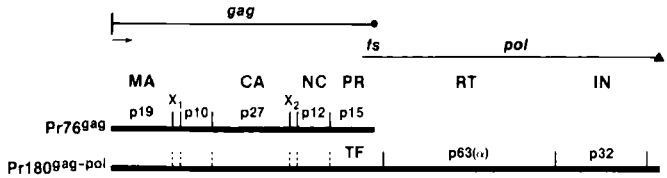
In-frame readthrough:

Mo-MuLV

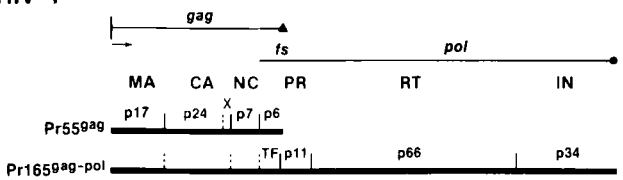


Single frameshift:

RSV

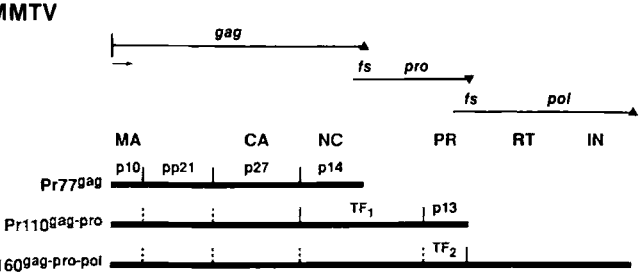


HIV-1

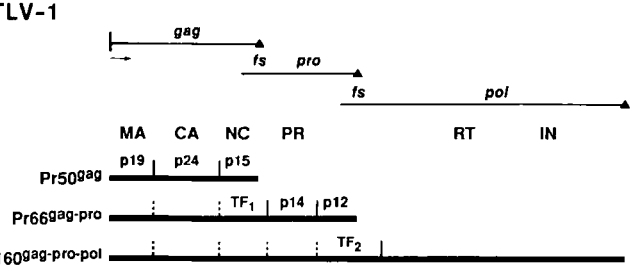


Double frameshift:

MMTV



HTLV-1



the end of the *gag* gene. Most of the ribosomes engaged in Gag protein synthesis terminate peptide chain elongation in response to this termination codon (as would be expected), resulting in the synthesis of the Gag structural polyprotein. However, a minority of these ribosomes engage in translational suppression: that is, they continue peptide synthesis *beyond* the termination codon, generating a large Gag–Pol fusion protein. Because the Gag polyprotein precursor normally performs the self-assembly processes responsible for virus assembly, it seems very likely that the Gag moiety of the Gag–Pol fusion protein participates in this self-assembly along with the authentic Gag polyprotein, so that the Gag–Pol fusion protein is incorporated into the virus particle. Thus the use of translational suppression in *pol* gene expression simultaneously modulates the relative level of Pol protein synthesis and provides for the inclusion of the Pol proteins in the virion.

Though all retroviruses utilize translational suppression in Pol protein synthesis, different groups of retroviruses use two completely distinct types of translational suppression. One of these is in-frame or readthrough suppression, and the other is ribosomal frameshifting.

In the viruses using readthrough suppression, the *gag* and *pol* coding sequences are in the same reading frame and are separated by a single UAG termination codon. A minority of the ribosomes engaged in Gag protein synthesis insert an amino acid in response to this UAG triplet, rather than terminating synthesis and releasing the product, and then continue beyond it to translate the *pol* sequences.

In contrast, in the viruses using ribosomal frameshifting, the *gag* and *pol* coding sequences are out of frame with respect to each other, with the *pol* sequences placed in the -1 position relative to the *gag* sequences. At some point near, but prior to the termination codon signaling the end of the *gag* gene, a minority of the ribosomes engaged in Gag protein synthesis translate a codon and insert the corresponding amino acid, but advance only two, rather than three, bases. Thus, this subpopulation of ribosomes shifts from the *gag* reading frame to the *pol* reading frame: this transition allows it to bypass the *gag* termination codon and synthesize a Gag–Pol fusion protein. (As will be discussed below, some yeast retrotransposons exhibit ribosomal frameshifting in the $+1$, rather than the -1, direction.)

The distinction between these two mechanisms should be emphasized. Ribosomal frameshifting may occur at a considerable distance on the mRNA from the *gag* termination codon, and eliminating this termination codon by mutation does not prevent frameshifting (Jacks *et al.*, 1988a). Thus, frameshifting is quite independent of the presence of the

termination codon. In contrast, readthrough suppression represents unusual behavior of ribosomes *at* a termination codon.

Some retroviruses, including mouse mammary tumor virus (MMTV), the mammalian type D viruses, and members of the human T cell leukemia virus (HTLV) group, actually use ribosomal frameshifting twice, rather than once, in the synthesis of the *pol* gene product: once between the *gag* and protease-coding sequences, and again between the protease gene (*pro*) and that for RT and IN (see Fig. 1). (Another deviation from the general schemes presented above should also be noted: in the avian type C and the foamy retroviruses, the PR is encoded on the 5′ side of the *gag* termination codon, so that it is encoded within the *gag* region rather than the *pol* region of the genome.)

The present review summarizes our present state of knowledge concerning translational suppression in retroviruses. Other viruses using similar mechanisms are mentioned only briefly and tangentially. For a description of the historical development of our understanding of this subject (as regards retroviruses), readers are referred to an excellent review by Jacks (1990).

II. Readthrough Suppression

As noted above, a single inframe UAG termination codon separates the *gag* and *pol* genes in the mammalian type C retroviruses (Shinnick *et al.*, 1981; Tamura, 1983; Herr, 1984; Etzerodt *et al.*, 1984; Kato *et al.*, 1987; Weaver *et al.*, 1990). The synthesis of a large Gag–Pol fusion protein in murine leukemia virus (MuLV)-infected cells, at a molar ratio of approximately 1:20 with that of the Gag polyprotein, was originally observed by Jamjoom *et al.* (1977), who suggested the existence of a translational control mechanism governing the synthesis of this product. As discussed in a recent review by Jacks (1990), an obvious alternative was the presence of a distinct *gag–pol* mRNA from which the termination codon in the viral genome had been removed by splicing.

Virtually all of the studies on the mechanism of Gag–Pol synthesis in mammalian type C retroviruses have been performed with MuLVs. One important experimental approach that has been used in analyzing the synthesis of the fusion protein has been *in vitro* translation (Jackson and Hunt, 1983). Early experiments (Kerr *et al.*, 1976; Murphy *et al.*, 1978) showed that translation reactions programmed with virion RNA were capable of synthesizing the Gag–Pol precursor as well as the

Gag polyprotein. Indeed, the ratio of Gag to Gag–Pol produced in these *in vitro* systems appeared to be comparable to that observed in the infected cell. Thus *gag–pol* mRNA is present in the virus particle. Because virion RNA is apparently a single, homogeneous species, the *gag–pol* mRNA is evidently indistinguishable from the genomic RNA of the virus in its approximate size and composition. However, these observations could not exclude the possibility that virions contained a second RNA species, distinct from the genomic RNA by virtue of the fact that the termination codon at the end of *gag* had been removed by splicing.

In a significant extension of these studies on *in vitro* translation of virion RNA, it was found that the relative level of synthesis of the Gag–Pol fusion protein could be increased by the addition of purified yeast amber suppressor tRNA (Philipson *et al.*, 1978; Murphy *et al.*, 1980). Although no sequence data were available at the time these experiments were performed, this finding strongly suggested that a translational suppression mechanism was responsible for the synthesis of the Gag–Pol fusion protein, and that a UAG codon was present between the *gag* and *pol* regions of the viral genome.

More recently, these experiments have been refined by using mRNA synthesized *in vitro* from an infectious clone of proviral DNA, rather than RNA isolated from virions, to direct the synthesis of both Gag and Gag–Pol polyproteins in rabbit reticulocyte lysates (Feng *et al.*, 1989a). In general, results obtained with this system reflected the earlier observations with viral RNA quite closely. The synthesis of Gag–Pol product in response to this synthetic, presumably completely homogeneous mRNA is obviously strong support for the idea that the two proteins are synthesized from the same template as a result of translational suppression.

Remarkably, the mechanism of synthesis of the Gag–Pol fusion protein was in large part clarified by a single, seminal observation. In 1985, Yoshinaka *et al.* reported the isolation and N-terminal amino acid sequence of the PR of Moloney murine leukemia virus (Mo-MuLV) (Yoshinaka *et al.*, 1985a). Because the PR gene is at the 5′ end of the *pol* coding region (Levin *et al.*, 1984; Crawford and Goff, 1985; Katoh *et al.*, 1985), it was expected that the protein would begin with an amino acid sequence encoded entirely 3′ of the *gag* termination codon. However, when the actual N-terminal sequence was compared with the MuLV proviral DNA sequence (Shinnick *et al.*, 1981), it was found that the first four residues represented the last four codons of the *gag* gene. They were followed by a glutamine residue, and then by the amino acids encoded at the 5′ end of the *pol* gene. It thus appears that the PR is

formed by cleavage of the Gag–Pol fusion protein at a site four residues before the last *gag*-encoded amino acid. Similar results were also obtained by sequencing the PR of feline leukemia virus (FeLV) (Yoshinaka *et al.*, 1985b). The protein sequences showed clearly that (1) PR is synthesized by translation of a mRNA containing the UAG termination codon present in the viral genome, rather than by a spliced mRNA lacking this codon, and (2) the UAG termination codon is suppressed *in vivo* by a glutamine tRNA.

A. Role of Context in Readthrough Suppression

1. Comparison of Cellular and Viral Termination Signals: Evidence for "Suppression Signal" in Viral mRNA

The fact that the UAG codon at the *gag–pol* junction of mammalian type C retroviral RNAs is suppressed at a significant level implies that normal cells of higher eukaryotes contain tRNAs capable of inserting an amino acid in response to this termination codon. However, the UAG termination codons found at the ends of many coding sequences in the cellular genome (Kohli and Grosjean, 1981; Brown *et al.*, 1990) are not suppressed significantly (Capone *et al.*, 1986; Sedivy *et al.*, 1987; Martin *et al.*, 1989). How can we explain this striking difference between translation of cellular and viral mRNAs? One simple explanation for the efficiency of termination at normal cellular termination sites would be that they are actually multiple, tandem termination codons. However, survey of a number of eukaryotic coding sequences showed that this is not the case (Kohli and Grosjean, 1981; Brown *et al.*, 1990). Another possibility is that normal termination codons are associated with signals favoring termination, in essence "protecting" them from the cellular tRNAs capable of suppression. Finally, the viral mRNA may contain signals promoting the suppression event. As discussed by Valle and Morch (1988) and below, it seems likely that both of these latter hypotheses are correct.

The nature of the difference between cellular and viral mRNAs was approached by the construction and analysis of nonsense mutants (mutants containing termination codons at internal positions) in reporter genes, including chloramphenicol acetyltransferase (Capone *et al.*, 1986; Martin *et al.*, 1989) and poliovirus (Sedivy *et al.*, 1987). These termination codons are thus at sites that, unlike the retroviral *gag–pol* junction, are not designed for efficient suppression, but are also not the location of natural termination codons. When these mutant genes were expressed in normal mammalian cells, it was found that they are not

suppressed to a detectable extent (Capone *et al.*, 1986; Sedivy *et al.*, 1987; Martin *et al.*, 1989). This observation suggested that suppression during translation of the viral mRNA occurs because this RNA contains positive signals favoring suppression.

The possibility that type C retroviral mRNA contains signals promoting suppression was tested directly in a series of experiments by Panganiban (1988). He isolated a restriction fragment of the MuLV genome containing 37 codons from the 3′ end of *gag*, the UAG codon at the *gag–pol* junction, and 62 codons from the 5′ end of *pol*. This fragment, corresponding to the region of MuLV mRNA near the *gag–pol* junction, was inserted into a construct at the 5′ end of the *lacZ* gene. When this construct was transfected into mammalian cells, a significant level of readthrough of the UAG codon was observed; indeed, the level of suppression (about 10%) was quantitatively comparable to that which actually takes place during translation of the viral genome *in vivo*. A control construct, containing an inframe UAG codon but lacking the viral sequences, showed no detectable suppression. In a somewhat analogous experiment, Honigman *et al.* (1991) changed a CAG codon within the Mo-MuLV *gag* gene (nt 1623–1625; Shinnick *et al.*, 1981) to UAG, and observed no suppression of this UAG codon in an *in vitro* translation system.

These results provided direct evidence that the viral mRNA contains signals promoting suppression at the *gag–pol* junction, and also implied that these signals do not extend beyond the limits of the restriction fragment used in Panganiban's experiments (Panganiban, 1988). In addition, because the constructs did not encode any viral proteins, and because essentially identical results were obtained in both virus-infected and uninfected cells (Panganiban, 1988), the data argue that the viral signals that result in suppression at the *gag–pol* junction are completely cis-acting.

On the other hand, a number of studies suggested that natural termination codons are, in fact, found in contexts that are unfavorable for suppression. Thus, when nonsense suppressor tRNAs (mutant tRNAs whose anticodons pair with termination codons, but which can be acylated with amino acids and function in translation) were introduced into higher eukaryotic cells together with the nonsense mutants of chloramphenicol acetyltransferase and poliovirus discussed above, the suppressor tRNAs could be shown to suppress these termination codons (Capone *et al.*, 1986; Sedivy *et al.*, 1987). However, the presence of these tRNAs had a surprisingly small effect on the pattern of cellular protein synthesis observed in two-dimensional electrophoresis (Bienz *et al.*, 1981). There was also very little effect on cell growth (Sedivy *et al.*,

1987). This observation, that the suppressor tRNAs did not detectably suppress many of the "natural" termination codons at the ends of cellular genes, strongly suggested that the latter codons are in contexts protecting them from suppression (or promoting efficient termination).

At present, there is little information on the nature of the signals discussed above. In prokaryotes, it has been shown that nonsense mutants of *lacI* (Miller and Albertini, 1983; Bossi, 1983) or other genes (Engelberg-Kulka, 1981) are more efficiently suppressed if the nucleotide immediately following the termination codon is a purine. Studies of this type have not, to our knowledge, been performed in eukaryotes.

When the sequence around natural termination codons in eukaryotic genes was analyzed, a very strong bias was found for purines at the position immediately 3′ of the termination codon (Kohli and Grosjean, 1981; Brown *et al.*, 1990). This bias was even more striking in genes expressed at a high level (Brown *et al.*, 1990). These observations might suggest that a purine at this position is an important element of the hypothetical signal promoting efficient termination at natural termination sites. However, the suppressible termination codon at the *gag–pol* junction of all known mammalian type C retroviruses is also followed by a G residue! Clearly, the viral signal promoting suppression must extend beyond this position.

2. *Possible Suppression Signals in Mammalian Type C Retroviral mRNA*

One approach that might point to signals favoring suppression is to compare sequences of different viruses that use readthrough suppression, to determine whether conserved sequences occur near the suppressible termination codon. Figure 2 presents the sequences of the 20 codons on either side of the gag termination codon of Mo-MuLV (Shinnick *et al.*, 1981), AKR MuLV (Herr, 1984), spleen necrosis virus (SNV) (Weaver *et al.*, 1990), and baboon endogenous virus (BaEV) (Kato *et al.*, 1987). (The latter viruses are much more distantly related to the two MuLVs than the two MuLVs are to each other.) Since all of these sequences were obtained from infectious molecular clones, they all represent portions of mRNAs which successfully engage in readthrough suppression.

Inspection of Fig. 2 shows that there is very limited sequence conservation [8 out of 60 nucleotides (nt), or 13%] on the 5′ side of the termination codon. However, there are several striking features on the 3′ side. These include a GG pair immediately beyond the UAG codon; a GU CAG GG sequence in the second, third, and fourth *pol* codons; a run of

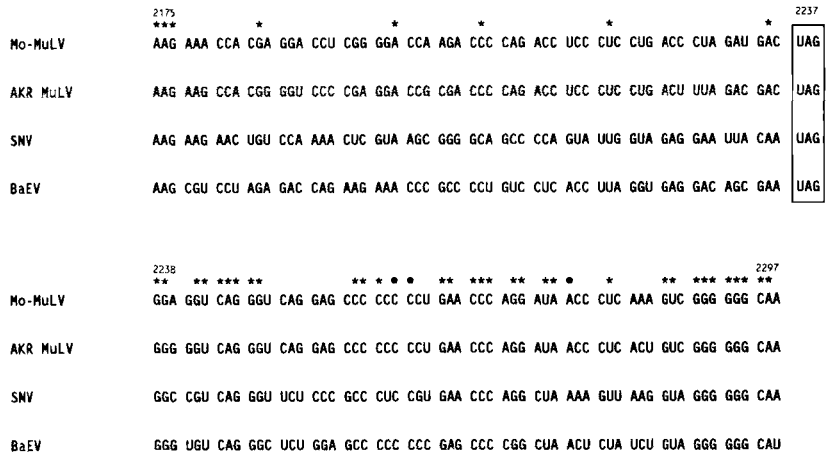


FIG. 2. Comparison of the nucleotide sequence around the UAG termination codon in Mo-MuLV (Shinnick *et al.*, 1981), AKR MuLV (Herr, 1984), spleen necrosis virus (SNV) (Weaver *et al.*, 1990), and baboon endogenous virus (BaEV) (Kato *et al.*, 1987). The nucleotide sequence 60 bases 5′ of the UAG termination codon and 60 bases 3′ of the UAG codon in Mo-MuLV, AKR MuLV, SNV, and BaEV is shown. Nucleotide positions are indicated for Mo-MuLV (Shinnick *et al.*, 1981). Nucleotides that are identical in all four viruses are denoted by an asterisk. The UAG codon present in each sequence is boxed.

six pyrimidines in the seventh, eighth, and ninth *pol* codons (all C residues except for one U in spleen necrosis virus); and a run of six G residues, followed by CA, in the eighteenth, nineteenth, and twentieth *pol* codons. There are also a number of conserved bases between the run of pyrimidines and that of the Gs. In all, nearly 60% of the bases in this 60-nt stretch are identical in the four viruses. The degree of conservation observed on the 3′ side of the *gag* termination codon is strongly suggestive of a possible role for these sequences in suppression.

One obvious possibility is that the signal for suppression is contained in secondary structures in the viral RNA, rather than in specific sequences; such structures clearly play a role in many instances of ribosomal frameshifting (see below). One candidate structure is a potential stem–loop in MuLV depicted in Fig. 3. To investigate this possibility, Jones *et al.* (1989) made point mutations in sequences surrounding the Mo-MuLV *gag–pol* junction that would destroy the stem and measured the effect on viral infectivity. Changes that would allow base pairing in the stem (Fig. 3), e.g., C2220 to U and G2252 to A, or C2220 to U alone, led to the production of infectious virions, whereas mutations that would destabilize the secondary structure, e.g., A 2223 to C and G2252 to A, or G2252 to A alone, did not. On the basis of these observations,

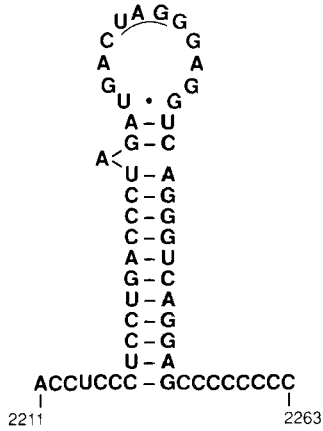


FIG. 3. Potential RNA secondary structure in Mo-MuLV RNA at the *gag–pol* junction. The stem–loop structure proposed by Shinnick *et al.* (1981) is shown. The structure within the nucleotide sequence 2211–2263 includes nucleotides 2217–2255; the UAG codon (underlined) is nucleotides 2235–2237. The nucleotide at position 2255 is indicated as a G rather than a C as originally reported by Shinnick *et al.* (1981). Miller and Verma (1984) showed that this change is one of two nucleotide changes that convert the noninfectious clone sequenced by Shinnick *et al.* (1981) to an infectious clone. A similar stem–loop structure for AKR MuLV was proposed by Herr (1984).

Jones *et al.* (1989) suggested that a region of secondary structure near the UAG codon must be preserved; however, their study did not test the effect of destabilizing mutations upstream of the UAG codon (e.g., A2223 to C alone) nor did they measure suppression directly.

In a related series of experiments carried out *in vitro*, Honigman *et al.* (1991) introduced destabilizing mutations into residues in the stem of the putative stem–loop structure (Fig. 3) at positions 5′ (nt 2222–2226, GACCC to AAUAU) and 3′ (nt 2246–2250, GGGUC to UCAUG) of the UAG codon. The upstream mutation had no effect on suppression in the *in vitro* system; in contrast, the downstream mutation prevented readthrough. These results led Honigman *et al.* (1991) to conclude that a secondary structure involving the UAG codon and nearby nucleotides at the MuLV *gag–pol* junction is unlikely to be important in suppression and in addition emphasized the role of the downstream sequences. It is interesting that stem–loop structures similar to the one shown in Fig. 3 probably do not exist in the viral mRNAs of other mammalian type C retroviruses that undergo readthrough suppression (Panganiban, 1988; ten Dam *et al.*, 1990). Even more importantly, it has been pointed out (Panganiban, 1988) that a stem–loop structure containing the UAG codon is unlikely, because the structure would

have to be disrupted before ribosome movement and misreading of the UAG codon could occur.

Computer analysis of the Mo-MuLV RNA sequence (ten Dam *et al.*, 1990) also raised the possibility that the *gag–pol* mRNA contains a pseudoknot structure (pseudoknots are stem–loop structures in which the bases in the loop are paired with bases downstream of the stem [Pleij *et al.*, 1985; Schimmel, 1989] see Section III). This structure might involve a long stretch of six C residues (nt 2256–2261; Shinnick *et al.*, 1981) beginning 19 nt downstream of the UAG codon, which could interact with a run of six G residues (nt 2289–2294) to form the second base-paired region of the pseudoknot. These runs of C and G residues are fairly well conserved in other mammalian type C retrovirus genomes (ten Dam *et al.*, 1990) (Fig. 2). In one study it was found that mutation of nucleotides in the first stem of a putative pseudoknot structure inhibited readthrough *in vitro*, but compensatory mutations did not restore activity (Honigman *et al.*, 1991).

Though particular structures have not yet been identified, it is clear that mutations in downstream sequences have an inhibitory effect on readthrough. In addition, it is noteworthy that mutation of a conserved sequence (Honigman *et al.*, 1991) (Fig. 2) immediately 3′ of the Mo-MuLV UAG termination codon, GGAG (nt 2238–2241) to ACGC, completely abolished *in vitro* synthesis of a Gag–Pol fusion protein (Honigman *et al.*, 1991). In summary, the mutational data, as well as the sequence conservation (Fig. 2), are all consistent with the possibility that, as in prokaryotes (Engelberg-Kulka, 1981; Miller and Albertini, 1983; Bossi, 1983), the suppression signal is contained within the downstream sequences.

The exact number of nucleotides required for readthrough in the MuLV system is under investigation. Based on an analysis similar to that shown in Fig. 2, Feng *et al.* (1990a) designed a miniconstruct of Mo-MuLV mRNA containing the last two codons of *gag*, the UAG termination codon at the *gag–pol* junction, and the first 19 codons of *pol*, and could show that the UAG codon was suppressed in rabbit reticulocyte lysates (Feng *et al.*, 1990a). This result suggested that a limited region of viral mRNA contains all the sequences needed for suppression and provided additional evidence that these sequences are largely, if not entirely, downstream sequences.

B. Alternate Stop Codons

The mutational analysis discussed above indicates that at least part of the signal that governs readthrough suppression is contained within

the primary sequence of the viral mRNA. An important question to consider is whether this signal is specific for the UAG codon or whether other termination codons can be substituted within the nucleotide context required for suppression. Feng *et al.* (1989b) used oligonucleotide-directed mutagenesis to change the UAG codon at the Mo-MuLV *gag–pol* junction to UAA or UGA. Both UAA and UGA were suppressed with the same efficiency as UAG in rabbit reticulocyte lysates (Feng *et al.*, 1989a,b). In the case of UAA, however, the system had to be supplemented with additional tRNA; rabbit liver tRNA or tRNA from uninfected or MuLV-infected NIH/3T3 cells were equally effective (Feng *et al.*, 1989b). This observation suggested that the tRNA that suppresses UAA is not abundant in the usual calf liver tRNA-supplemented rabbit reticulocyte lysate (Jackson and Hunt, 1983; Feng *et al.*, 1989a) and that the UAA suppressor tRNA is not unique to mouse cells. *In vivo* experiments carried out by transfecting intact viral genomes with UAA or UGA instead of UAG into Chinese hamster ovary (CHO) cells (conditions that do not permit selection of revertants to wild type) led to production of infectious virions with approximately the same titer as wild type (Feng *et al.*, 1989b). Similarly, the viral capsid (CA) protein and the Pol proteins, RT and IN, were present in equivalent amounts in virions derived from mutant and wild-type genomes (Feng *et al.*, 1989b). Indeed, the Gag and Gag–Pol precursor proteins were synthesized to the same extent in the cells transfected with UAG-, UAA-, or UGA-containing viral genomes (A. Rein, unpublished observations 1989). From these results, Feng *et al.* (1989b) concluded that (1) the signal(s) for UAG suppression are effective with UAA and UGA; (2) UAA is not an absolute termination codon in higher eukaryotes, as had been previously thought (Geller and Rich, 1980; Valle and Morch, 1988); and (3) mammalian cells and cell extracts contain tRNAs capable of suppressing UGA and UAA termination codons that appear in a retroviral context. Jones *et al.* (1989) also reported that mutant Mo-MuLV viral genomes with alternate stop codons can give rise to infectious virus particles.

The discovery that all three termination codons are suppressible in the MuLV system raised the possibility of identifying previously unknown suppressor tRNAs. The approach used was to translate a miniconstruct mRNA that has a short leader sequence containing an AUG codon, followed by two codons from the 3′ end of *gag*, a termination or sense codon, 19 codons from the 5′ end of *pol*, and the binding domain of protein A. The N-terminal amino acid sequence of the product was then determined by the Edman degradation technique (Feng *et al.*, 1990a). The predicted amino acid sequence of the first 20 amino acids of the

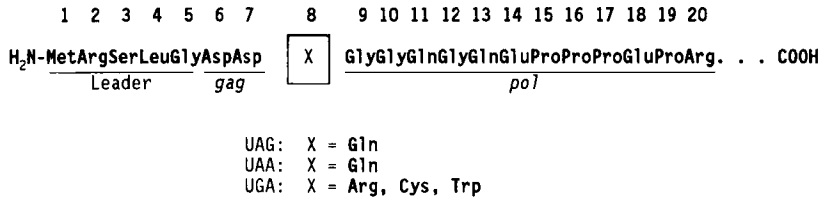


FIG. 4. Amino acid sequence at the Mo-MuLV readthrough site in a protein A fusion protein synthesized *in vitro*. The predicted amino acid sequence of the first 20 amino acids is shown. The boxed X at position 8 represents the amino acids inserted in response to either the UAG, UAA, or UGA termination codon.

fusion product is shown in Fig. 4; position 8 represents the residue at the Gag–Pol junction. Comparison of the Edman degradation patterns obtained using mRNAs with UAG and CAG (a glutamine code word) showed that the relative amounts of radioactive glutamine incorporated at the readthrough site and at two positions in Pol (Fig. 4) were the same for both products. This result demonstrated that UAG is translated predominantly, if not exclusively, as glutamine and pointed to the parallel between suppression *in vitro* and *in vivo*, where the UAG termination codon is known to be read as glutamine (Yoshinaka *et al.*, 1985a). In addition, sequence analysis of the UAA fusion protein showed that UAA, like CAG and UAG, directs the sole incorporation of glutamine at the Gag–Pol junction (Feng *et al.*, 1990a). This finding represented the first (and to date only) identification of an amino acid inserted in response to UAA in a higher eukaryote. Surprisingly, in the case of UGA, three amino acids, arginine, cysteine, and tryptophan, were inserted at the GAG–Pol junction (Feng *et al.*, 1990a). It is of interest that early codon recognition studies (Marshall *et al.*, 1967; Caskey *et al.*, 1968; Hatfield, 1972) indicated that arginine, cysteine, and tryptophan tRNAs can respond to UGA in the ribosomal binding assay of Nirenberg and Leder (1964). However, misreading of UGA as cysteine and arginine during protein synthesis had not been previously described for higher eukaryotes.

The rabbit reticulocyte lysate system used for this study (Feng *et al.*, 1989a) presumably contains other known mammalian suppressor tRNAs, i.e., two leucine UAG suppressors (Valle *et al.*, 1987) and a UGA suppressor that is acylated with serine (Hatfield, 1972; Hatfield *et al.*, 1982a) and ultimately converted *in vivo* to selenocysteyl-tRNA (Lee *et al.*, 1989). Despite the presumed presence of these suppressor tRNAs, leucine and serine were not inserted at the Gag–Pol junction and were

detected only at the predicted positions (Fig. 4) in the fusion products (Feng *et al.*, 1990a). This strict specificity exhibited by the MuLV suppression system raises the intriguing possibility that nucleotide context not only affects the efficiency of suppression, but also determines which tRNAs will function in suppression.

C. Suppressor tRNAs

Whereas cis-acting viral sequences clearly play a significant role in MuLV readthrough suppression, nonviral trans-acting factors are also of major importance. These factors presumably include (1) the single mammalian release factor that mediates termination in response to all three termination codons (Konecki *et al.*, 1977); (2) normal cellular tRNAs that act as suppressor tRNAs by misreading termination codons; and (3) other factors, such as ribosomal proteins or RNA sequences, which may have functional significance, but have not yet been identified in mammalian systems. The molecular mechanisms underlying the interaction between these factors are poorly understood. Presumably, there is a competition between suppressor tRNA and a release factor that determines whether suppression can take place. In addition, it has been proposed that suppression may be promoted by base pairing between one or two nucleotides immediately downstream of the termination codon in the message and the corresponding bases 5′ of the anticodon in the tRNA (Engelberg-Kulka, 1981; Panganiban, 1988).

Although it has been known for some time that a glutamine tRNA suppresses the UAG codon at the MuLV and FeLV *gag–pol* junctions (Yoshinaka *et al.*, 1985a,b), the glutamine isoacceptor that mediates this suppression has not been identified. Kuchino *et al.* (1987) sequenced two glutamine tRNAs from mouse liver: (1) a major species with the anticodon CUG and (2) a minor species (only 1–2% of the major species) having the anticodon U_mUG. The sequences of these glutamine tRNAs were the same except for the 5′ position of the anticodon and the nucleotides at positions 4 and 68 in the acceptor stem; both tRNAs occurred in hypo- and hypermodified forms. Both forms of the minor tRNA species were able to weakly suppress the tobacco mosaic virus (TMV) UAG codon in an *in vitro* suppression assay, but neither form of the major species had any suppressor activity (Kuchino *et al.*, 1987). Interestingly, misreading of a UAG termination codon by the minor glutamine isoacceptor would involve unusual codon–anticodon recognition by mispairing at both the first and third positions of the codon. Because the MuLV and TMV systems are not identical, it is not clear that the minor glutamine tRNA functions as the suppressor in MuLV

infection as proposed (Kuchino *et al.*, 1987). Thus, in contrast to the situation with MuLV (Feng *et al.*, 1990a), the tRNA specificity for *in vitro* suppression of the TMV UAG codon is less stringent and several tRNAs exhibit suppressor activity, including glutamine (Kuchino *et al.*, 1987) and leucine (Valle *et al.*, 1987) as well as the hypomodified tyrosine tRNA, which is the TMV suppressor *in vivo* (Bienz and Kubli, 1981; Beier *et al.*, 1984a).

Several groups have considered the question of whether virus infection affects the suppressor tRNA population. Kuchino *et al.* (1987) reported that infection with Mo-MuLV increased the amount of the minor glutamine tRNA. Other investigators have obtained different results. Feng *et al.* (1989a) found that infection with Mo-MuLV did not change the chromatographic profile of glutamine isoacceptors or the level of glutamine acceptor activity. Moreover, it could be shown that suppression of the Mo-MuLV UAG termination codon *in vitro* was stimulated to the same extent by tRNA isolated from MuLV-infected or uninfected NIH/3T3 cells (Feng *et al.*, 1989a). Similarly, as noted above, Panganiban (1988) observed that *in vivo* suppression of the UAG codon at the *gag–pol* junction occurred with the same efficiency in Mo-MuLV-infected and uninfected mouse cells. Taken together, these findings led Feng *et al.* (1989a) to conclude that the glutamine suppressor tRNA occurs normally within the tRNA population of uninfected cells and is not altered or induced in response to virus infection. In addition, the observation that all three termination codons can be suppressed with the same efficiency *in vitro* and *in vivo* (Feng *et al.*, 1989b) and the fact that several distinct suppressor tRNAs can function within the MuLV context (Feng *et al.*, 1990a; see below) are difficult to reconcile with a requirement for viral induction of suppressor tRNA.

The observation that a glutamine residue was inserted *in vitro* in response to a UAA termination codon at the MuLV *gag–pol* junction (Feng *et al.*, 1990a) clearly indicated that a glutamine tRNA mediates suppression of UAA in mammalian cells. As in the case of UAG, it is not known which isoacceptor functions as the UAA suppressor in the MuLV system. Suppression of both UAA and UAG termination codons by glutamine tRNAs has a precedent in yeast. The yeast glutamine tRNA, which can suppress UAA, normally decodes CAA (Pure *et al.*, 1985), whereas a different isoacceptor, which normally recognizes CAG, suppresses UAG (Weiss and Friedberg 1986; Lin *et al.*, 1986; W. A. Weiss *et al.*, 1987). Whether UAA and UAG are suppressed by two distinct glutamine tRNAs in mammalian cells as they are in yeast is not known. In this connection, it may be relevant that additional

tRNA must be added to mammalian extracts for efficient suppression of UAA, but not of UAG (Feng *et al.*, 1989b, 1990a).

Because the UGA termination codon at the *gag–pol* junction was decoded as three amino acids, arginine, cysteine, and tryptophan, a number of different tRNAs must mediate UGA suppression. Tryptophan tRNA involvement in UGA suppression has already been observed in normal mammalian cells and bacteria. Geller and Rich (1980) proposed that mammalian tryptophan tRNA can function as a UGA suppressor based on their finding that partially purified tryptophan tRNA from reticulocyte lysates stimulates *in vitro* suppression of a UGA termination codon in β-hemoglobin mRNA. In bacteria, wild-type tryptophan tRNA and a mutant tryptophan suppressor tRNA with a G-to-A change at position 24 (Hirsh, 1971) decode the UGG tryptophan codon and UGA *in vitro* (Hirsh and Gold, 1971) and *in vivo* (Raftery *et al.*, 1984). Interestingly, Buckingham and Kurland (1977) found that the suppressor tRNA also decodes the UGU cysteine codon with low efficiency *in vitro*. Because mammalian tryptophan tRNA, like its bacterial counterpart (Hirsh, 1971), is expected to have a CCA anticodon, interaction with UGA may require C–A mispairing at the third position of the codon. Similarly, insertion of cysteine (UGU and UGC codons) in response to UGA would be expected to involve mispairing at the third position of the codon. Although arginine has six codons and several isoacceptors (Hatfield, 1972), the most likely candidate for suppressor activity in this case would appear to be a CGA-decoding tRNA, which could suppress UGA by G–U mispairing at the first position of the codon, in analogy to the interactions of glutamine tRNAs with UAA and UAG in yeast (Pure *et al.*, 1985; Weiss and Friedberg, 1986; Lin *et al.*, 1986; W. A. Weiss *et al.*, 1987) and possibly in MuLV. The possibility that as of yet unidentified specialized suppressor tRNA(s) are involved in readthrough suppression at the MuLV *gag–pol* junction should also be considered. The subject of suppressor tRNAs in readthrough suppression in higher eukaryotes has also been reviewed by Valle (1989), Hatfield *et al.* (1990a,b), and Valle and Haenni (1991).

The work cited in this section shows that mammalian cells contain either four or five distinct species (depending on whether the same glutamine tRNA is used in UAG and UAA suppression) that can suppress termination codons at the MuLV *gag–pol* junction. As of yet none of these tRNAs has been definitively identified or characterized, but this will be important for future studies on the mechanism of readthrough suppression. For example, mutational analysis of tRNA structure as well as mRNA context should provide insights into the nature of the interactions between cis- and trans-acting factors.

D. Readthrough Suppression in Other Viruses

Though the present discussion has focused on readthrough suppression in retroviruses, it should be noted that this mechanism is also used by other viruses to modulate the level of synthesis of fusion proteins. Thus, in several alphaviruses, including Sindbis virus, a single UGA codon separates two open reading frames (Strauss *et al.*, 1983, 1984, 1988). It has been shown that this UGA codon is suppressed *in vivo* (Li and Rice, 1989). A number of plant viruses, including TMV (Pelham, 1978; Goelet *et al.*, 1982), carnation mottle virus (Guilley *et al.*, 1985), and beet necrotic yellow vein virus (Bouzoubaa *et al*, 1986), and use readthrough of a UAG codon. An elegant analysis by Skuzeski *et al.* (1991) has shown that in the case of TMV, the signal responsible for readthrough suppression is confined to the two codons immediately 3′ of the termination codon. It is intriguing to note that many plant viruses exhibiting readthrough suppression have a nearly identical sequence in this position, whereas others, such as carnation mottle virus, have a different sequence (Valle, 1989).

In both TMV (Ishikawa *et al.*, 1986) and Sindbis virus (Li and Rice, 1989), as in MuLV (Feng *et al.*, 1989b), readthrough occurs with each of the three possible termination codons. Despite this similarity in the different viral systems, sequence comparison shows no obvious homology in the sequences surrounding the suppressible termination codon (Feng *et al.*, 1990b).

III. Ribosomal Frameshifting

As noted in Section I, ribosomal frameshifting alters the reading frame of mRNA during translation, resulting in the expression of a single protein from two or more overlapping genes. Ribosomal frameshifting may operate in one of two directions, altering the reading frame in either the 5′ or 3′ direction. This phenomenon is well known in bacteria and has been reviewed elsewhere (Dayhuff *et al.*, 1986; Craigen and Caskey, 1987; R. B. Weiss *et al.*, 1987; R. Weiss *et al.*, 1988; Hughes *et al.*, 1989; Atkins *et al.*, 1990; Murgola, 1990). In eukaryotes, a shift in the reading frame in the 3′ direction has been described thus far only in yeast, whereas that in the 5′ direction has been described in yeast, plants, and animals. For example, the retrovirus-like retrotransposon, Ty, and the double-stranded RNA viruslike particle, L-A, both contain two large overlapping reading frames that are aligned differently in yeast (for review see Wickner,

1989). In Ty, the different reading frames are aligned by a frameshift of one nucleotide in the 3′ (or +1) direction (Wilson *et al.*, 1986; Clare *et al.*, 1988), whereas in L-A they are aligned by a frameshift of one nucleotide in the 5′ (or −1) direction (Icho and Wickner, 1989; Dinman *et al.*, 1991). In higher eukaryotes, ribosomal frameshifting occurs or is suspected of occurring in the −1 direction in a number of mammalian and avian retroviruses, in the avian infectious bronchitis virus (IBV) (Brierley *et al.*, 1987, 1989), in certain plant viruses (Miller *et al.*, 1988; Xiong and Lommel, 1989), in transposable elements in *Drosophila*, and in the mouse intracisternal A-particle (mouse IAP) (for reviews see Jacks, 1990; Hatfield *et al.*, 1990a,b; Hatfield and Oroszlan, 1990). In fact, the *gag* and *pol* genes of most vertebrate retroviruses occur in different reading frames and ribosomal frameshifting in the −1 direction is required to align the overlapping frames. Interestingly, as noted above, some of these retroviruses require two frameshift events, one between *gag–pro* and one between *pro–pol*, to express the Gag–Pro–Pol fusion protein (see Fig. 1).

The means of unequivocally demonstrating ribosomal frameshifting is to sequence the transframe protein (i.e., the protein that spans the overlapping reading frame) through the frameshift site and compare the resulting peptide to the corresponding RNA (template) sequence (Hizi *et al.*, 1987; Jacks *et al.*, 1988a,b; Weiss *et al.*, 1989; Nam *et al.*, 1992). The fact that viral genes may be tandem, lie in different reading frames, and appear to be overlapping does not necessarily mean that they are expressed by ribosomal frameshifting, even though the gene organization may be analogous to that of other genetic systems utilizing frameshifting, and, for that matter, even though such genes are expressed as a fusion protein. For example, the cauliflower mosaic virus capsid protein and RT genes are tandem and lie in different reading frames, but RT is expressed separately from the capsid protein (see Schultze *et al.*, 1990; Wurch *et al.*, 1991, and references therein). In addition, in the hepatitis B virus, the *X* and *C* genes are expressed as a fusion protein (where these genes may occur in different reading frames), but recent evidence suggests that this fusion protein is not synthesized by ribosomal frameshifting (see Lo *et al.*, 1990, and references therein).

In the present review, we have included those genetic systems in which ribosomal frameshifting has unequivocally been shown to occur, or is suspected of occurring based on the presence of an established frameshift signal within the overlapping region (see Table I). We examine ribosomal frameshifting, in both the −1 and +1 directions, in detail below. For comparison, several other, nonretroviral systems will also be briefly considered.

A. Frameshifting in the −1 direction

Frameshifting in the −1 direction in eukaryotes was first demonstrated by Jacks and Varmus (1985), who reported that both the Gag protein and the Gag–Pol fusion protein of Rous sarcoma virus (RSV) could be synthesized in rabbit reticulocyte lysates programmed with a single species of RNA encoding the RSV *gag* gene and an adjacent portion of the downstream *pol* gene. The fact that both polypeptides were formed from a single species of RNA in approximately the same ratios as found *in vivo* provided strong evidence that ribosomal frameshifting accounted for the alignment of the different reading frames in RSV. Ribosomal frameshifting was unequivocally demonstrated when the *in vivo*-made transframe protein spanning the *gag–pro* overlap of MMTV was sequenced and found to contain amino acid residues that matched the corresponding nucleotide template, except for a shift by one nucleotide in the −1 direction (Hizi *et al.*, 1987). A detailed examination of the frameshift site, of information encoded in viral RNA for frameshifting, of possible models for frameshifting, of unique features of the frameshift site, and of the possible role of tRNA in frameshifting are presented below.

TABLE I

DETERMINED AND PROPOSED RIBOSOMAL FRAMESHIFT SITES AND SIGNALS IN VIRUSES AND IN TRANSPOSABLE ELEMENTS OF EUKARYOTES

Source[a]	Overlap	Bases in overlap window	Distance from frameshift site to 3′ end of overlap	Bases at and around the frameshift signal[b]	Class and subclass[c]
MMTV	*gag–pro*	16	3	UCA AAA AAC UUG	AAAC-1
BLV	*gag–pro*	49	0	UCA AAA AAC UAA	AAAC-1
HTLV-1, STLV-1	*gag–pro*	37	18	CCA AAA AAC UCC	AAAC-1
HTLV-2	*gag–pro*	28	18	GAA AAA AAC UCC	AAAC-1
EIAV	*gag–pol*	241	195	CCA AAA AAC GGG	AAAC-1
HTLV-1	*pro–pol*	178	156	CCU UUA AAC CAG	AAAC-2
STLV-1	*pro–pol*	121	99	CCU UUA AAC CGG	AAAC-2
HTLV-2	*pro–pol*	373	18	CCU UUA AAC CUG	AAAC-2
BLV	*pro–pol*	22	0	CCU UUA AAC UAG	AAAC-2
MHV	1a–1b	76	18	UUU UUA AAC GGG	AAAC-2
IBV	F1–F2	40	30	UAU UUA AAC GGG	AAAC-2
BEV	1a–1b	10	3	GAU UUA AAC UGU	AAAC-2
SRV-1	*gag–pro*	181	147	CAG GGA AAC GGA	AAAC-3
SRV-2, MPMV	*gag–pro*	181	147	CAG GGA AAC GGG	AAAC-3
Visna	*gag–pol*	124	45	CAG GGA AAC AAC	AAAC-3

TABLE I *(continued)*

Source[a]	Overlap	Bases in overlap window	Distance from frameshift site to 3′ end of overlap	Bases at and around the frameshift signal[b]	Class and subclass[c]
Mouse IAP	*gag–pol*	34	3	CUG GGU UUU**C**CU	UUUU-1
BYDV	*p39–pol*	13	0	GUG GGU UUU**U**AG	UUUU-1
SRV-1, MPMV	*pro–pol*	22	0	GGA AAU UUU**U**AA	UUUU-2
SRV-2	*pro–pol*	22	0	GGA AAU UUU**U**AG	UUUU-2
17.6	*gag–pol*	46	30	GAA AAU UUU**C**AG	UUUU-2
RCNMV	*p27–pol*	7	0	GAG GAU UUU**U**AG	UUUU-3
HIV-1	*gag–pol*	241	234	AAU UUU UUA**G**GG	UUUA-1
HIV-2	*gag–pol*	283	267	GGU UUU UUA**G**GA	UUUA-1
SIV	*gag–pol*	343	213	GGU UUU UUA**G**GC	UUUA-1
gypsy	*gag–pol*	70	51	AAU UUU UUA**G**GG	UUUA-1
RSV	*gag–pol*	58	0	ACA AAU UUA**U**AG	UUUA-2
MMTV	*pro–pol*	13	0	CAG GAU UUA UGA	UUUA-3
L-A	*gag–pol*	130	96	CAG GGU UUA GGA	UUUA-4

[a] Abbreviations not given in text: STLV-1, simian leukemia T cell virus, type 1; EIAV, equine infectious anemia virus; MHV, mouse hepatitis virus; BEV, Berne virus; MPMV, Mason–Pfizer monkey virus; BYDV, barley yellow dwarf virus; SIV, simian immunodeficiency virus. References: MMTV (Hizi *et al.*, 1987; Jacks *et al.*, 1987; Moore *et al.*, 1987); BLV (Rice *et al.*, 1985; Sagata *et al.*, 1985); HTLV-1 (Seiki *et al.*, 1983; Hiramatsu *et al.*, 1987; Inoue *et al.*, 1986); STLV-1 (Inoue *et al.*, 1986); HTLV-2 (Mador *et al.*, 1989; Shimotohno *et al.*, 1985); EIAV (Stephens *et al.*, 1986; Kawakami *et al.*, 1987); MHV (Lee *et al.*, 1991); IBV (Brierley *et al.*, 1987); BEV (Snijder *et al.*, 1990); SRV-1 (Power *et al.*, 1986); SRV-2 (Thayer *et al.*, 1987); MPMV (Sonigo *et al.*, 1986); Visna (Sonigo *et al.*, 1985); mouse IAP (Mietz *et al.*, 1987); BYDV (Miller *et al.*, 1988); 17.6 (a transposable element in *Drosophila*) (Saigo *et al.*, 1984); RCNMV (Xiong and Lommel, 1989); HIV-1 (Jacks *et al.*, 1988b; Ratner *et al.*, 1985; Wain-Hobson *et al.*, 1985; Sanchez-Pescador *et al.*, 1985); HIV-2 (Guyader *et al.*, 1987); SIV (Franchini *et al.*, 1987; Chakrabarti *et al.*, 1987); *gypsy* (a transposable element in *Drosophila*) (Marlor *et al.*, 1986); RSV (Hughes *et al.*, 1989; Jacks *et al.*, 1988a; Schwartz *et al.*, 1983); L-A (a double-stranded RNA viruslike particle in yeast) (Dinman *et al.*, 1991).

[b] Underlined bases designate heptanucleotide sequences within the overlaps that are associated or are suspected of being associated with frameshifting (see text and Jacks *et al.*, 1988a; Jacks, 1990; Hatfield *et al.*, 1990a,b). The bold letter at the 3′ end of the frameshift signal designates that the precise site of the frameshift has been established by sequencing the transframe peptide of one member of the subclass (see text and footnote *c*).

[c] Frameshift signals are placed into classes on the basis of the consensus sequence (i.e., AAAC, UUUU, and UUUA) and into subclasses on the basis of the upstream triplet such that members of subclasses have identical heptanucleotide signals. The fact that the transframe peptide has been sequenced from one member of the subclass, establishing the precise site of the frameshift (see text), demonstrates that the precise site for each member is known.

1. Identifying the Frameshift Site

The boundaries of the overlapping region (also called the overlap or frameshift window) are set by the termination codon in the 0 frame (e.g., the termination codon at the end of *gag*) and the nearest upstream termination codon in the −1 frame. The frameshift must occur of course within the overlapping region. The size of the overlap can vary from seven nucleotides [as observed in the frameshift window of red clover necrotic mosaic virus (RCNMV)] (Xiong and Lommel, 1989; see below) to several hundred nucleotides in length (as observed in the *pro–pol* frameshift window of HTLV-2 (Shimotohno *et al.*, 1985). The site of the frameshift may occur anywhere within the overlap from the extreme 3′ end to the 5′ end.

A search of the overlapping regions within the *gag–pol* genes of retroviruses and the equivalent regions in other viruses and retroelements of higher organisms, including those in *Drosophila* and in the mouse IAP, each of which requires −1 frameshifting for alignment of different reading frames, reveals the occurrence of one of three common consensus sequences. As shown in Table I, the sequences are A AAC, U UUA, or U UUU, where AAC, UUA, and UUU decode asparagine, leucine, and phenylalanine, respectively, in the 0 frame (see also Jacks *et al.*, 1988a). It is of interest to note that IBV (Brierley *et al.*, 1987) and two plant viruses, barley yellow dwarf virus (BYDV) (Miller *et al.*, 1988) and RCNMV (Xiong and Lommel, 1989), each contain one of the common sequences within the respective overlap shared by two different reading frames. As will be discussed in a later section, the signal for frameshifting that encompasses the common consensus sequence is actually a heptanucleotide, as shown in Table I.

The frameshift site is the 3′ base at the end of the consensus sequence. This was most clearly demonstrated by sequencing the transframe peptide generated *in vitro* from the RSV sequence, UUAUA (Jacks *et al.*, 1988a), where UUA is a leucine codon that is read in the *gag* frame and is part of the U UUA consensus sequence, and the *pro–pol* sequence of HTLV-1, AACCA (Nam *et al.*, 1991), where AAC is an asparagine codon that is read in the *pro* frame and is part of the A AAC consensus sequence (see Table I). As shown in Fig. 5, leucine and isoleucine were generated from the UUAUA RSV sequence, and thus the UUA codon was read as leucine in the *gag* (or 0) frame and the AUA codon was read as isoleucine in the *pol* (or −1) frame. Likewise, asparagine and proline were generated from the HTLV-1 *pro–pol* AACCA sequence and thus AAC was decoded as asparagine in the *pro* frame, and CCA was decoded as proline in the *pol* frame. These studies show

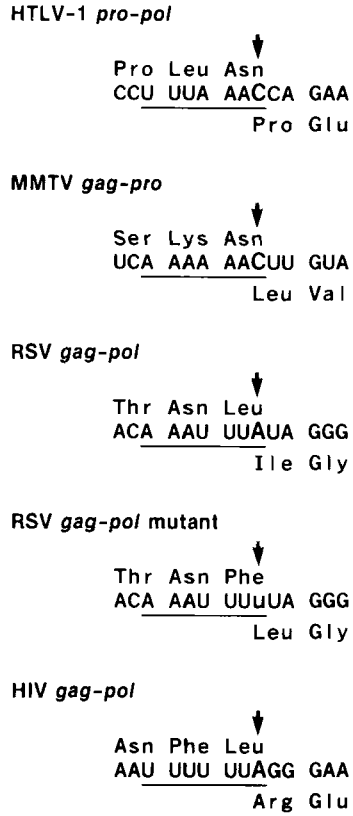


FIG. 5. Nucleotide and amino acid sequences of the transframe proteins at the frameshift site in selected retroviruses. Underlined nucleotides designate the frameshift signals. The amino acid sequences above the nucleotide sequences are translated from the 0 reading frame and those below are translated from the −1 reading frame. The arrows point to the frameshift site; u is the site of the mutation in Rous sarcoma virus (RSV) (see text).

unequivocally that the site of the frameshift is the 3′ end of the consensus sequence. It should be noted that frameshifting has not yet been shown to occur at the 3′ end of the U UUU common sequence in any naturally occurring (wild-type) overlap. However, it seems likely that the 3′ U in U UUU is a frameshift site, because changing the RSV consensus sequence, U UUA, to U UUU promotes frameshifting and incorporates phenylalanine into the transframe peptide *in vitro* (Jacks *et al.*, 1988a).

The transframe protein had been sequenced from other retroviruses in studies performed earlier than those described above. For sequencing, the MMTV Gag–Pro transframe protein was purified from virus (Hizi *et al.*, 1987) and the human immunodeficiency virus (HIV) peptide was synthesized *in vitro* from the *gag–pol* construct (Jacks *et al.*, 1988b). In MMTV, the sequence AAA AAC UUG UAA occurs at the 3′ end of *gag* where A AAC is the consensus sequence (Table I and Fig. 5). The transframe protein generated from this sequence contained lysine and asparagine (decoded by AAA and AAC, respectively, in the 0 frame), followed by leucine (decoded by either UUG in the 0 frame or CUU in the −1 frame) and then valine (decoded by GUA in the −1 frame) (Hizi *et al.*, 1987), but due to overlapping Leu codons in the 0 and −1 frames, the precise site of the frameshift could not be determined. However, by analogy to the results described above with HTLV-1, it seems likely that the frameshift occurred at the 3′ end of the AAC codon within the MMTV *gag–pro* overlap as shown in Fig. 5.

In sequencing the transframe peptide generated *in vitro* from HIV-1, Jacks *et al.* (1988b) observed that the shift occurred at a leucine residue corresponding to the Leu codon, UUA, which is part of the consensus sequence U UUA (Table I and Fig. 5). However, both leucine and phenylalanine were present at the position of the frameshift (in a ratio of 7:3), making assignment of the precise site uncertain. It is of interest to note that Weiss *et al.* (1989) translated the HIV frameshift signal in *Escherichia coli* cells, sequenced the resulting transframe peptide, and observed both leucine and phenylalanine at the frameshift site in a ratio of about 3:1. A similar analysis of the transframe peptide generated from the MMTV *gag–pro* frameshift signal, A AAA AAC, in *E. coli* yielded asparagine and lysine at the frameshift site in a ratio of about 7:3 (Weiss *et al.*, 1989). These investigators proposed that a major shift occurred at the normal frameshift site (i.e., on the UUA codon in HIV and on the AAC codon in MMTV), whereas a minor shift occurred at the immediate upstream codon (i.e., on the UUU codon in HIV-1 and on the AAA codon in MMTV) (see Fig. 5). Kingsman *et al.* (1990) have also considered the possibility that the HIV-1 U UUU UUA frameshift sequence is quite slippery (Wilson *et al.*, 1988), such that a minor shift that occurs at the upstream UUU codon would account for the occurrence of two amino acids at the frameshift site. Other possibilities that may account for the occurrence of two amino acids at the HIV frameshift site as observed by Jacks *et al.* (1988b) also warrant consideration. For example, Jacks *et al.* (1988b) proposed that a portion of the Leu-tRNA that is decoded by UUA at the ribosomal A-site in the 0 frame may come off the ribosome after the slip to the −1 reading frame. This

event would expose the UUU codon, which then would be decoded by Phe-tRNA. Another possibility is that the frameshift site, which most certainly has unique features that make it slippery, may be more favorable to misreading such that Phe-tRNA misreads the leucine codon in HIV-1 and Lys-tRNA misreads the asparagine codon in MMTV prior to the frameshift (see further discussion on misreading within the frameshift signal in Section III,A,5).

The studies discussed above were carried out in heterologous systems (Jacks *et al.*, 1988b; Weiss *et al.*, 1989). However, it is important to know whether two amino acids occur at the frameshift site in the naturally occurring transframe proteins (i.e., in the Gag–Pol fusion protein of HIV-1 and the Gag–Pro fusion protein of MMTV) synthesized under normal physiological conditions in virus-producing cells. A transframe peptide derived from the naturally occurring HIV Gag–Pol fusion protein was sequenced (eight amino acids in length) and was found to contain Phe–Leu–Arg (L. Henderson, personal communication 1991), where Phe–Leu is read in the *gag* frame (decoded by UUU UU**A**) and Arg is read in the *pol* frame (decoded by **A**GG, where **A** is the site of the frameshift) (see Fig. 5). However, this study did not fully exclude the possibility that a second transframe peptide with the sequence Phe–Phe–Arg may also exist in HIV-1. Similarly, amino acid sequencing of the naturally occurring transframe protein of MMTV did not indicate microheterogeneity (Hizi *et al.*, 1987), but the possibility cannot be excluded here either. Therefore, additional studies will be required to establish the very important point of whether amino acid sequence heterogeneity exists at the frameshift site in the naturally occurring transframe proteins or peptides.

2. *Mechanism of the Alignment of Reading Frames*

As discussed above, the sequence of various transframe peptides showed that the frameshift occurred at the 3′ end of the consensus sequences, as shown in Table I. Although these studies pinpoint the exact site of the frameshift, they do not determine how the different reading frames are aligned, i.e., whether alignment occurs by overlapping reading (where the base at the 3′ end of the consensus sequence would be read twice) or by doublet decoding (where only two of three bases of the 0 frame codon within the consensus sequence would be read). However, an experiment by Jacks *et al.* (1988a) in which a single base change at the 3′ end of the frameshift signal results in two new amino acids in the transframe peptide demonstrates that the alignment occurs by overlapping reading. Jacks *et al.* (1988a) changed the RSV UU**A***UA* sequence, which codes for leucine in the 0 frame and isoleu-

cine in the −1 frame, to UUU*UA*, where U is the altered base at the 3′ end of the frameshift signal (see Fig. 5). The resulting transframe peptide generated from this sequence contained phenylalanine, which was decoded by UUU, and leucine, which was decoded by U*UA* (see Fig. 5). The base at the 3′ end of the frameshift signal is therefore read twice, once in the 0 frame and once in the −1 frame; thus, the alignment of the different reading frames occurs by overlapping reading as originally proposed by Hizi *et al.* (1987; see also Hatfield and Oroszlan, 1990).

3. Identifying Information Encoded in RNA for Frameshifting (Cis-Acting Sequences)

Two different kinds of information have been identified in viral RNA that have a role in signaling the frameshift event: (1) as noted above, a heptanucleotide sequence that encompasses the frameshift site on its 3′ end and the immediate six upstream bases (Table I) (Jacks *et al.*, 1988a; Wilson *et al.*, 1988; Brierley *et al.*, 1989; Dinman *et al.*, 1991; Nam *et al.*, 1992), and (2) RNA secondary structure, which occurs just downstream of the heptanucleotide sequence (Jacks *et al.*, 1987, 1988a; Brierley *et al.*, 1989; Weiss *et al.*, 1989; Dinman *et al.*, 1991). Both types of information, which have been termed cis-acting sequences (for review see Jacks, 1990), are further examined below.

a. Information at the Frameshift Site. The heptanucleotide sequence that signals the frameshift event (see Table I) was identified largely by site-directed mutagenesis studies within and/or around the frameshift region (Jacks *et al.*, 1988a,b; Nam *et al.*, 1988; Wilson *et al.*, 1988; Brierley *et al.*, 1989; Dinman *et al.*, 1991; Nam *et al.*, 1992). In RSV, alteration of the U UU bases within the consensus U UUA sequence (where UUA is read in the 0 frame and A is the site of the frameshift) inhibited ribosomal frameshifting most severely (Jacks *et al.*, 1988a). Alteration of any of the three bases immediately upstream of the RSV consensus sequence (i.e., A AA within A AAU UUA) reduced the level of frameshifting, but only moderately compared to that observed with the U UU bases.

A similar observation was made in the double-stranded yeast virus, L-A, in which changes in the first three bases of the heptanucleotide frameshift signal (i.e., in G GG of the G GGU UUA frameshift signal) had a smaller inhibitory effect on frameshifting than those occurring in the downstream U UU sequence (Dinman *et al.*, 1991). Alteration of the first triplet to any identical three bases (i.e., G GG to C CC, A AA, or U UU) within the L-A frameshift signal maintained efficient frameshifting. The changes to pyrimidines in the first triplet resulted in higher levels of frameshifting than the corresponding purines, with U

UU giving the highest level. The latter observations suggest that the homopolymeric U sequences may provide an extremely slippery signal (see also Weiss *et al.*, 1989; Kingsman *et al.*, 1990; and Section III,A,3,b).

Frameshift signals in the HTLV-1 *gag–pro* and *pro–pol* overlaps have also been identified by mutagenesis studies. Alteration of the *gag–pro* A AAA AAC sequence to A AUA UUC inhibited frameshifting (Nam *et al.*, 1988), as did alteration of the *pro–pol* U UUA AAC sequence to U UUA AGC (Nam *et al.*, 1991). Deletion of the U triplet in the *pro–pol* frameshift signal also inhibited frameshifting, but changing the sequence immediately downstream of the AAC codon from CAGAA to UGCAG did not affect the frameshift event (Nam *et al.*, 1991).

Mutation of the 3′ terminal A in the sequence U UUU UUA, where A is at the frameshift site, to any of the other three bases did not inhibit the level of frameshifting in RSV or HIV (Jacks *et al.*, 1988a; Wilson *et al.*, 1988). In L-A, changing the 3′ terminal A in UUA to U or C likewise did not inhibit frameshifting, but changing this site to G did reduce the level of frameshifting by 5- to 10-fold (Dinman *et al.*, 1991). The frameshift event was enhanced slightly by changing this A to U (note that this change results in a new consensus sequence, U UUU) in all three viruses examined (Jacks *et al.*, 1988a; Wilson *et al.*, 1988; Weiss *et al.*, 1989; Dinman *et al.*, 1991). The fact that altering the U UUA sequence in L-A to A AAC maintained wild-type frameshift levels demonstrates that the same set of consensus signals that have been observed to be associated with −1 frameshifting in higher eukaryotes (i.e., U UUA, U UUU, and A AAC) also can work efficiently in yeast (Dinman *et al.*, 1991).

In contrast to the results described above in which mutants at the 3′ end of the homopolymeric U sequence had only moderate effects on the level of ribosomal frameshifting in mammalian protein synthesis, changing the base at the 3′ end of the homopolymeric A frameshift signal had far more pronounced effects (Chamorro *et al.*, 1992). Alteration of the C in the MMTV *gag–pro* frameshift (A AAA AAC) signal to U reduced the level of frameshifting severalfold, whereas changing this base to A or G was even more inhibitory, with the A AAA AAG sequence exhibiting the most severe inhibition. In contrast, changing the 3′ terminal C in the MMTV *gag–pro* (homopolymeric A) frameshift signal to A or G increased, and did not decrease, the level of frameshifting in *E. coli* (Weiss *et al.*, 1989). The level of frameshifting increased about 1.5 times with a 3′ terminal A and about 30 times with a 3′ terminal G.

In mammalian cells, AAC and AAU are decoded by the same isoacceptor and in *E. coli* the same tRNA decodes AAA and AAG. Presumably, in each case the two codons are normally translated with roughly equal efficiencies by the corresponding cognate isoacceptors. It is therefore of considerable interest that the levels of frameshifting are quite different with AAC compared to AAU in mammalian cells, as are the levels in *E. coli* with AAG compared with AAA. The implications of these observations are further considered in Section III, A,5.

One major point to be emphasized in the present discussion is that the frameshift signal that encompasses the frameshift site is a heptanucleotide sequence, which was identified largely by mutagenesis studies (Jacks *et al.*. 1988a,b; Wilson *et al.*, 1988; Brierley *et al.*, 1989; Dinman *et al.*, 1991). This conclusion is further supported by the observation that efficient frameshifting occurs when the heptanucleotide frameshift signal in IBV is altered such that it is flanked on both sides by a termination codon [i.e., when the codon immediately upstream in the −1 frame and that immediately downstream in the 0 frame of the heptanucleotide signal are stop codons (S. Inglis, personal communication 1989)]. Furthermore, the frameshift window in RCNMV is only seven nucleotides in length (Xiong and Lommel, 1989).

b. Information Downstream of the Frameshift Site. Not all of the information necessary for frameshifting is located within the heptanucleotide sequence shown in Table I. Initially, Jacks *et al.* (1987) demonstrated that the ability of the *gag–pro* and *pro–pol* overlaps of MMTV to promote frameshifting was lost when these regions were inserted into another genetic context. Thus, these two overlaps (which contain the heptanucleotide frameshift signal) are not sufficient in themselves to carry out frameshifting. Jacks *et al.* (1987) also suggested that potential stem–loop structures that occur just downstream of the overlaps in MMTV may provide the additional information required for the frameshift event (see above). The presence of a potentially stable RNA secondary structure just downstream of homopolymeric A sequences in the *gag–pro* overlaps of bovine leukemia virus (BLV) and HTLV-1 was first noted by Rice *et al.* (1985). Potential stem–loop structures occur just downstream of the frameshift sites or suspected sites in each overlapping region sequenced to date from higher eukaryotes (actually within 9 to 10 bases of the frameshift site in all cases; see Jacks, 1990; Kingsman *et al.*, 1990). Interestingly, in a survey of several hundred bases on either side of the frameshift site, Le *et al.* (1989) found that the most thermodynamically stable secondary structure was always a stem–loop located within several bases downstream of the frameshift site. Although the stem–loop structures vary considerably

in size of both the stem and the loop, the type of configuration that can be generated, in base composition, and in stability, the striking conservation of their position suggests that they may have an important role in frameshifting. Several examples of possible stem–loop structures are shown in Fig. 6.

Efficient frameshifting has been shown to be dependent on the presence of a stem–loop structure that occurs just downstream of the frameshift site in RSV (Jacks *et al.*, 1988a), in IBV (Brierley *et al.*, 1989), in L-A (Dinman *et al.*,1991), and in the *pro–pol* overlap of HTLV-1 (Nam *et al.*, 1991). Jacks *et al.* (1988a) used deletion–substitution and site-specific mutations to show a direct relationship between the presence of a stem–loop structure and the efficiency of frameshifting in RSV. Deletion of bases within the stem–loop structure virtually abolished frameshifting. Disruption of base pairings within the stem by generating specific stem-destabilizing mutations resulted in a decrease in frameshifting, whereas restoring these base pairings by generating specific stem-restabilizing mutations rescued frameshifting (Jacks *et al.*, 1988a).

Site-directed alteration of a number of specific bases further downstream of the stem–loop structure in IBV inhibited frameshifting (Brierley *et al.*, 1989). The latter bases are complementary to those in the loop of the stem–loop, which raises the possibility that many of the downstream bases interact with the loop in IBV, resulting in a tertiary structure known as a pseudoknot. The role of the stem–loop structure and pseudoknot in IBV frameshifting has been examined in further detail (Brierley *et al.*, 1991). For efficient frameshifting in IBV, the two pseudoknot stems must be in close proximity to each other and must be essentially intact. However, small changes in the loops of the pseudoknot did not affect frameshifting. These investigators also observed that the pseudoknot could not be replaced by a simple stem–loop structure of similar overall size and composition; thus the pseudoknot conformation is a requirement for frameshifting.

In the double–stranded RNA viruslike particle, L-A, evidence has been presented that the stem–loop structure that is immediately downstream of the frameshift site exists as part of a pseudoknot, and the entire pseudoknot structure is required for efficient frameshifting (Dinman *et al.*, 1991). Furthermore, the stem–loop structures that occur immediately downstream of the frameshift signals or proposed frameshift signals in a number of retroviruses (ten Dam *et al.*, 1990), as well as in murine coronavirus gene 1 (Lee *et al.*, 1991), are capable of forming a pseudoknot. In MMTV, the presence of a pseudoknot downstream of the *gag–pro* frameshift signal is required for optimal frame-

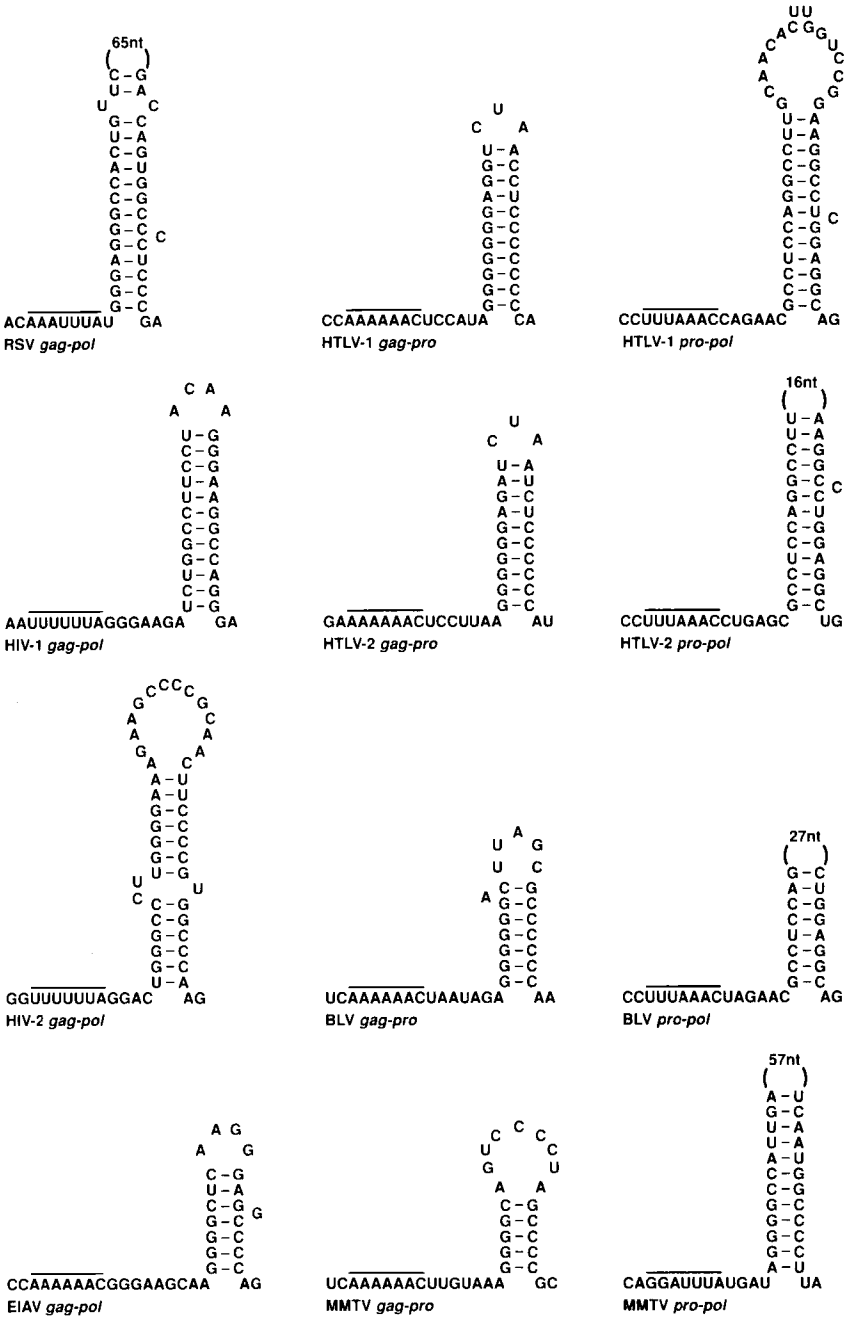


FIG. 6. Potential stem–loop structures that occur just downstream of the frameshift site in a number of retroviruses.

shifting (Chamorro *et al.*, 1992). The occurrence and role of pseudoknots in retroviruses and in other RNA structures have been reviewed elsewhere (Schimmel, 1989; Wyatt *et al.*, 1989; Pleij, 1990; ten Dam *et al.*, 1990).

It is of interest to note that the position of the stem–loop structure relative to the frameshift site is critical for efficient frameshifting, as has been carefully demonstrated for IBV (Brierley *et al.*, 1989). Altering the distance between the stem–loop structure and the frameshift site by as few as three bases in either direction inhibits frameshifting in IBV. Thus, it appears that the ribosomal frameshift event, at least in some retroviruses, requires a carefully positioned downstream stem–loop structure that may, as in the case of IBV (Brierley *et al.*, 1989), exist as part of a pseudoknot.

The role of the downstream RNA secondary and/or tertiary structures may be to impede translation at the frameshift site long enough for the shift to occur (Rice *et al.*, 1985; Jacks *et al.*, 1987; Brierley *et al.*, 1989; Weiss *et al.*, 1989; Atkins *et al.*, 1990; Jacks, 1990; Kingsman *et al.*, 1990). As noted above, the stability of the stem–loop structures, their sizes, and the type of configurations that can be generated from them vary considerably. It remains to be determined how the variation in the stem–loop structure influences the efficiency of frameshifting.

The initial experiments involving the role of the potential stem–loop structure in the frameshift event in HIV suggested that this structure was not required for efficient frameshifting. For example, Madhani *et al.* (1988) tested a variety of mutations encompassing the potential stem–loop downstream of the HIV frameshift site, and these mutations (with one exception) had no effect on frameshifting. Wilson *et al.* (1988) focused on the ability of a short oligonucleotide encoding the heptanucleotide U UUU UUA frameshift signal in HIV to carry out frameshifting. They observed that this sequence worked as efficiently *in vitro* (in rabbit reticulocyte lysates) as *in vivo* (in yeast cells) whether or not the downstream stem–loop structure was present. However, it should be noted that these experiments were not carried out under normal physiological conditions for HIV *gag–pol* expression. Recent *in vivo* studies (carried out in mammalian cells) provide evidence that the downstream stem–loop structure in HIV is required for optimal frameshifting (H. Varmus, personal communication 1991), demonstrating that this structure indeed has an important role for efficient synthesis of the HIV Gag–Pol fusion protein.

It should also be noted that the experiments which show that HIV frameshifts equally well with and without a downstream stem–loop structure (Jacks *et al.*, 1988b; Wilson *et al.*, 1988) suggest that this

frameshift signal is more slippery than those examined in other genetic systems in which the role of the downstream stem–loop has been shown (under similar assay conditions) to be required for efficient frameshifting (Jacks *et al.*, 1987; Jacks *et al.*, 1988a; Brierley *et al.*, 1989). These observations led Kingsman *et al.* (1990) to speculate that the homopolymeric A and U frameshift sequences (i.e., A AAA AA and U UUU UU) may be more slippery than other frameshift signals, and therefore that their requirements for a downstream stem–loop structure to aid the frameshift event may be less stringent. However, in eukaryotes the requirements for efficient frameshifting with a perfect homopolymeric A (i.e., all of the bases are As) sequence apparently are different from that of a perfect homopolymeric U sequence. Mutation of the 3′ terminal C in the A AAA AAC frameshift signal to A (or to G) severely inhibits frameshifting (Chamorro *et al.*, 1992). In addition, Dinman *et al.* (1991) have observed that frameshifting in yeast was more efficient when the L-A frameshift signal contained homopolymeric pyrimidine sequences than when it contained homopolymeric purine sequences. Interestingly, in the study by Dinman *et al.* (1991), the most efficient frameshift signal contained six tandem U bases.

Translation of the MMTV *gag–pro* and HIV frameshift signals with and without the corresponding downstream stem–loop structure in *E. coli* cells shows that the presence of this structure has only a slight to moderate effect on enhancing frameshifting in this heterologous system (Weiss *et al.*, 1989). The transframe peptide generated from the MMTV *gag–pro* frameshift signal (A AAA AAC) in *E. coli* with and without the downstream stem–loop structure was sequenced (Weiss *et al.*, 1989). Interestingly, asparagine occurred predominantly at the frameshift site with the intact stem (in an approximate ratio of 3:1 with lysine), whereas in the absence of the stem–loop, lysine occurred predominantly at the frameshift site (in an approximate ratio of 2:1 with asparagine). One interpretation of these results is that the stem specifically enhances frameshifting on the AAC codon, with a minor frameshift occurring on the upstream AAA codon, whereas in the absence of the stem, the major shift occurs on the first slippery codon, AAA, with a minor shift on the downstream AAC codon (Weiss *et al.*, 1989). This observation is further considered below (see Section III,A,5).

4. Simultaneous-Slippage Model for Ribosomal Frameshifting

During the frameshift event the aminoacyl-tRNA, which is located at the ribosomal A-site, and the peptidyl-tRNA, which is located at the

ribosomal P-site, are translocated by one nucleotide in the 5′ direction. It is not entirely clear how frameshifting is accomplished, but Jacks *et al.* (1988a) have proposed that it occurs by simultaneous slippage of both the aminoacyl-tRNA and the peptidyl-tRNA by one nucleotide in the −1 direction, resulting in both tRNAs decoding a new set of codons (see Fig. 7A). Following slippage, the ribosome is prepared to read the −1 frame; normal transfer of the peptidyl-tRNA to the aminoacyl-tRNA and its translocation to the P-site bring the first codon in the −1 frame to the A-site. Then, normal decoding of the A-site and transfer of the nascent peptide to the incoming aminoacyl-tRNA consummate reading in the −1 frame (Fig. 7A).

Site-directed mutagenesis studies that show that all seven bases within the heptanucleotide frameshift signal are essential to efficient frameshifting (Jacks *et al.*, 1988a; Wilson *et al.*, 1988; Brierley *et al.*, 1989; Dinman *et al.*, 1991) and sequence analyses of the transframe peptide (Hizi *et al.*, 1987; Jacks *et al.*, 1988a,b; Weiss *et al.*, 1989; Nam *et al.*, 1991) that show that the shift to the −1 reading frame occurs at the ribosomal A-site (Jacks, 1990) provide support for the simultaneous-slippage model. These studies strongly suggest that the six bases within the 0 frame of the signal span both the ribosomal A- and P-sites and the six bases within the −1 frame (following the slippage) also span the A- and P-sites. After the slip to the −1 frame has taken place, the aminoacyl-tRNA and peptidyl-tRNA decode a new set of codons; the base sequence within the frameshift signal is such that a minimal amount of mismatching occurs (see Fig. 8 and Section III,A,5).

Weiss *et al.* (1989) have proposed a somewhat different simultaneous-slippage model for frameshifting based on their observations on translation of the retroviral homopolymeric A and U frameshift signals in *E. coli*. The major difference between the Jacks and Varmus model (for review see Jacks, 1990) and that of Weiss *et al.* is that the latter takes into account the probable presence of three sites on the ribosome, the aminoacyl-tRNA (A), the peptidyl (P), and the exit (E) sites, and the possibility that the shift occurs after peptide bond formation (see Fig. 7B). The Jacks and Varmus model proposes that the frameshift occurs before peptide bond formation while the ribosome is stationary. Weiss *et al.* (1989) suggested that the downstream stem–loop structure may exert its influence during translocation of the 3′ codon (within the frameshift signal) from the A- to P-site and the slip may occur while the tRNAs exist, albeit transiently, as hybrids in the E-/P- and P-/A-sites immediately after peptide bond formation (see Fig. 7B and the legend for details).

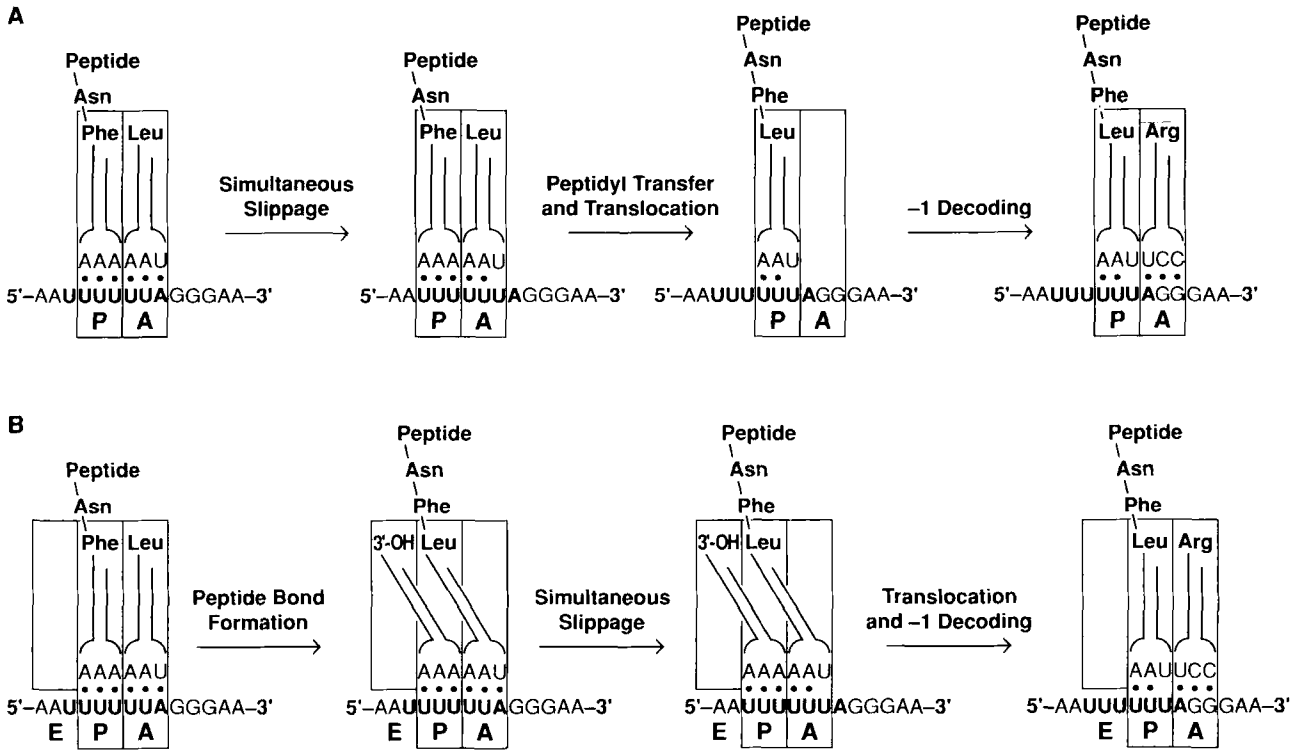


FIG. 7. Simultaneous-slippage models for ribosomal frameshifting in the −1 direction. (A) The simultaneous-slippage model of Jacks and Varmus (Jacks *et al.*, 1988a; Jacks, 1990) (see text for details). (B) The simultaneous-slippage model of Weiss *et al.* (1989). This model takes into account the occurrence of three ribosomal frameshift sites (A, P, and E) and suggests that the shift occurs after peptide bond formation (see text). Interestingly, the tRNA may exist (albeit transiently) as a hybrid occupying simultaneously the E/P- and P-/A-sites. This model also takes into account that the growing polypeptide remains stationary at the P-site (see Weiss *et al.*, 1989, for details and additional references).

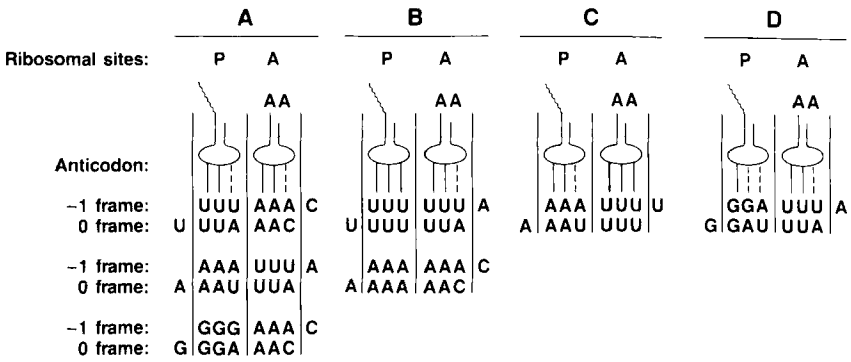


FIG. 8. Codon–anticodon interactions within frameshift signals at the ribosomal A- and P-sites before and after the shift of the reading frame. Codon sequences represent a summary of the ribosomal frameshift signals determined in vertebrate retroviruses (see Table I). The signals are arranged into four classes (columns A–D) depending on the codon–anticodon interaction after the frameshift event as follows: Shift from the 0 to the −1 frame results in misreading, or reading two out of three bases in both the ribosomal A- and P-sites (A), just the A-site (B), or just the P-site (C); in column D, a shift to the −1 frame results in reading only one base in the P-site and two bases in the A-site by the standard Watson–Crick base pairings. Squiggly lines signify the nascent polypeptides attached to tRNAs in the P-sites; AA represents the amino acid attached to tRNA in the A-site, and the dashed lines represent mismatching in codon–anticodon interactions between standard Watson–Crick base pairs.

5. *Role of tRNA in Frameshifting (Trans-Acting Factors) and Unique Features of the Frameshift Site*

Jacks *et al.* (1988a) noted that only A AAC, U UUU, and U UUA occur within the frameshift signals of retroviruses and other genetic elements in eukaryotes (see Table I), in which AAC, UUU, and UUA are decoded at the ribosomal A-site in the 0 frame. The observation that only three codons occurred at the ribosomal A-site and the finding that altering the consensus sequence within the frameshift signal to A AAA or G GGG inhibits frameshifting led these investigators to propose the existence of specialized "shifty" tRNAs that promote frameshifting (Jacks *et al.*, 1988a). These tRNAs and the corresponding codons are $tRNA^{Asn}$ (AAC), $tRNA^{Phe}$ (UUU), and $tRNA^{Leu}$ (UUA). These tRNAs are characterized by the fact that $tRNA^{Asn}$ contains the highly modified queuine (Q) base in the 5′ position of its anticodon, $tRNA^{Phe}$ contains the highly modified wyebutoxine (Wye) base in the 3′ position next to its anticodon (see Hatfield *et al.*, 1990b, and references therein), and

$tRNA^{Leu}$ lacks a highly modified base in its anticodon loop (Valle *et al.*, 1987).

The status of the tRNAs utilized in and around the frameshift signals in cells infected with HIV-1, BLV, HTLV-1, and simian retrovirus-1 (SRV-1) has been examined (Hatfield *et al.*, 1989, and unpublished observations). Interestingly, most of the Phe-tRNA from HIV- and SRV-1-infected cells lacked the highly modified Wye base in its anticodon loop, and most of the Asn-tRNA from BLV-, HTLV-1-, and SRV-1-infected cells lacked the highly modified Q base in its anticodon loop. Thus, a correlation exists between the occurrence of hypomodified Asn-tRNA and Phe-tRNA in retrovirus-infected cells and their utilization in translating codons within the respective frameshift signals (see Table I).

The appearance of hypomodified isoacceptors in retrovirus-infected cells, most certainly, is not a virally encoded phenomenon. Thus, a question may be raised as to the cause of hypomodification of specific isoacceptors in the host tRNA population after viral infection. This question has been addressed previously by Katze and collaborators (1983), who considered a number of possibilities to explain a deprivation of Q base in tRNA in tumor cells. It should be noted that Q base is obtained in the diet of mammals and is inserted in tRNA by an enzyme designated as queuine tRNA-ribosyltransferase (tRNA-GRT). These investigators suggested that a deficiency of Q base in tumor cell tRNA occurs in part because the requirements for this base exceed the dietary intake. This may be due to an increase in tRNA turnover and growth rate, to an inefficient salvage pathway, and/or to the possible occurrence of inhibitors of tRNA-GRT. Of these possibilities, it is tempting to speculate that in retrovirus-infected cells, the metabolism of the host may be altered in response to viral infection such that the host produces a new metabolite or more of a given metabolite, which acts as an inhibitor of an enzyme involved in production of the hypermodified base within tRNA. For example, 7-methylguanine and pteridine occur in mammalian cells and these metabolites are inhibitors of tRNA-GRT (see Katze *et al.*, 1983; French *et al.*, 1991, and references therein).

What is the possible role of hypomodified isoacceptors in ribosomal frameshifting? Clearly, the lack of a hypermodified base in the anticodon loop of tRNA would create more space in and around the frameshift site; in turn, this might facilitate frameshifting by allowing greater flexibility of movement of the anticodon (Hatfield *et al.*, 1989; Hatfield and Oroszlan, 1990). It is of interest to note that the presence of modified bases within the anticodon loop of tRNA restricts wobble, whereas their absence expands the decoding potential (Randerath *et al.*, 1979;

Bienz and Kubli, 1981; Beier *et al.*, 1984a,b; Björk *et al.*, 1987, 1989; Wilson and Roe, 1989; Claesson *et al.*, 1990). More specifically, among these studies it has been shown that some tRNAs lacking a modified base in their anticodon promote frameshifting (Björk *et al.*, 1989), whereas others promote misreading (Randerath *et al.*, 1979; Bienz and Kubli, 1981; Beier *et al.*, 1984a,b; Björk *et al.*, 1987; Wilson and Roe, 1989; Claesson *et al.*, 1990). In addition, with respect to tRNAs normally containing a Q or Wye base, the coding properties of tRNAs lacking Q base (Bienz and Kubli, 1981; Beier *et al.*, 1984a,b; Björk *et al.*, 1987; Meier *et al.*, 1985) and Wye base (Smith and Hatfield, 1986) differ from those of the corresponding fully modified isoacceptors. In light of these studies, it is tempting to speculate that the "shifty" tRNAs that promote frameshifting are hypomodified isoacceptors.

In the simultaneous-slippage model of frameshifting, the tRNAs involved must have a dual function. First, they must promote frameshifting, and then after the shift to the −1 frame, they must misread or read only two out of three bases of the new set of codons (see Fig. 8). Interestingly, the frameshift signals are designed to minimize misreading after the frameshift event. Within the heptanucleotide frameshift signals involved in a shift to the −1 reading frame (see Table I), the bases in the first two positions of the downstream codon (i.e., UU, AA, or GG) are identical to the base in the 3′ position of the upstream codon (i.e., U, A, or G, respectively). The shift to the new reading frame, therefore, maintains similar codon–anticodon interactions provided the tRNAs in the ribosomal A- and P-sites misread the base in the 3′ position of the −1 frame codon or read only two out of three bases. Codon–anticodon complexes within the various frameshift signals in retroviruses sequenced to date are summarized in Fig. 8. The only exceptions to the presence of identical bases in the first two positions of the downstream codon and the terminal position of the upstream codon in the frameshift signals shown in Table I are in that of RCNMV and the *pro–pol* signal of MMTV. These two heptanucleotide signals contain an Asp codon, GAU, and a shift to the −1 reading frame results in Asp tRNA decoding a glycine codon, GGA. The shift onto the GGA codon requires that mismatching takes place between the first and second positions of the Asp tRNA anticodon and the middle and third positions of the corresponding codon as shown in Fig. 8 D.

The frameshift site manifests several unique features. For example, only three naturally occurring codons (UUA, UUU, and AAC) have been found to occupy this site (Jacks *et al.*, 1988a; Jacks, 1990). Moreover, the site and the heptanucleotide sequence that encompasses the frameshift site constitute a slippery region such that the reading

frame of the corresponding mRNA may be altered (Jacks *et al.*, 1988a; Wilson *et al.*, 1988; Weiss *et al.*, 1989; Jacks, 1990; Kingsman *et al.*, 1990). An additional feature of the frameshift site, in contrast to what occurs in normal translation, is that the same isoacceptor may decode one cognate codon more efficiently than another at this site. That is, $tRNA^{Phe}$ decodes UUU slightly more efficiently than UUC (Jacks *et al.*, 1988a; Wilson *et al.*, 1988; Weiss *et al.*, 1989; Dinman *et al.*, 1991) and $tRNA^{Asn}$ decodes AAC several times more efficiently than AAU (Chamorro *et al.*, 1992). In *E. coli* the same $tRNA^{Lys}$ decodes AAG at the mutant MMTV *gag–pol* frameshift site about 20 times more efficiently than AAA (Weis *et al.*, 1989). Clearly, the frameshift site involves a form of misreading because the same isoacceptor presumably decodes these cognate codons with similar efficiencies at other mRNA sites.

Meier *et al.* (1985) have shown that in the absence of Q base (G is in the wobble position of the anticodon), $tRNA^{His}$ shows a strong preference for CAC codons, whereas $tRNA^{His}$ with Q base shows a slight preference for CAU codons. This study provides a model for Q-deficient tRNAs preferentially reading XAC codons, as is found in decoding the asparagine AAC/AAU codon set at the MMTV *gag–pro* wild-type and mutant frameshift sites (see above). Thus, it is possible that the frameshift site has a specific requirement for a hypomodified isoacceptor that can preferentially decode one of its cognate codons. In addition, because the frameshift site invokes misreading (at least among cognate tRNA codons), then in light of the observation that two amino acids may occur at the HIV *gag–pol* (Jacks *et al.*, 1988b; Weiss *et al.*, 1989) and MMTV *gag–pro* (Weiss *et al.*, 1989) frameshift sites, a question can be raised whether the occurrence of two amino acids is caused by (1) frameshifting at separate sites (as suggested in Weiss *et al.*, 1989) or (2) a high level of misreading at a single frameshift site as a result of using heterologous systems (see also Section III,B,2).

B. Frameshifting in the +1 Direction

The only examples observed thus far in eukaryotes of frameshifting in the +1 direction are in the Ty elements that occur in the yeast, *Saccharomyces cerevisiae*. The Ty elements are a family of retrotransposons that are about 5.5 kilobases in length and are flanked by direct repeats of 330–340 bp designated delta sequences in Ty1 and Ty2 and sigma sequences in Ty3. The Ty elements replicate through a DNA intermediate in a fashion similar to retrovirus replication and Ty occurs within viruslike particles designated Ty-VLPs. The Ty proteins of

the VLP coat are encoded within *TYA*, which is analogous to the retroviral *gag* gene: the Ty PR, RT, and IN are encoded within *TYB*, which is analogous to the retroviral *pol* gene (for reviews see Wickner, 1989; Garfinkel, 1991). *TYA* and *TYB* overlap each other by 38–44 bp and the *TYB* reading frame is offset from *TYA* by one base in the 3′ direction. *TYB* is expressed as a fusion protein with the *TYA* gene product, and expression of the fusion protein requires ribosomal frameshifting in the +1 direction (Clare and Farabaugh, 1985; Mellor *et al.*, 1985; Wilson *et al.*, 1986; Clare *et al.*, 1988; Belcourt and Farabaugh, 1990; S. Sandmeyer, personal communication 1991). Thus, expression of *TYA–TYB* is like that observed in retroviruses with the exception that circumventing the termination codon at the end of the *TYA (gag)* gene is effected by ribosomal frameshifting in the +1 instead of the −1 direction.

1. *Identifying Information Encoded in RNA for Frameshifting (the Frameshift Signal)*

Deletion and site-directed mutagenesis studies of bases within a 14-oligonucleotide sequence, which was previously shown to promote frameshifting in Tyl (Clare *et al.*, 1988), identified seven bases that are responsible for the frameshift event (Belcourt and Farabaugh, 1990). The seven bases are CUU AGG C; these codons are in the 0 reading frame. The studies demonstrated that all the information necessary for altering the reading frame in the +1 direction in Tyl is present in the seven-nucleotide sequence that constitutes the frameshift signal. Ty3 also frameshifts in the +1 direction and the signal for this event has been shown to exist somewhere within a 21-bp region of the 38-bp overlap; but, interestingly, this 21-bp region does not contain the seven-base signal used by Tyl (S. Sandmeyer, personal communication 1991).

2. *Identifying the Frameshift Site*

Belcourt and Farabaugh (1990) prepared a construct with the frameshift signal for Tyl 15 nucleotides (five codons) downstream of an initiation codon. The sequence of the transframe peptide generated from this construct showed that Leu–Gly, but not Leu–Arg, was decoded by the CUU AGG U frameshift signal, where **A** denotes the site of the frameshift and Leu–Gly are decoded by CUU**A** and GGU, respectively. Interestingly, the tetramer CUU**A** contains overlapping leucine codons in the 0 (CUU) and +1 (UU**A**) reading frames (Belcourt and Farabaugh, 1990), and yeast cells contain a leucine tRNA capable of decoding all six leucine codons (Weissenbach *et al.*, 1977). Thus, the frameshift in Tyl involves a slippage from one leucine codon in the 0 reading frame to an

overlapping leucine codon in the +1 reading frame (Belcourt and Farabaugh, 1990; see also below) and the site of the frameshift is the 3′ base of the downstream overlapping leucine codon.

3. *Model for +1 Frameshifting, the Role of tRNA, and Other Features of This Event*

The heptanucleotide frameshift signal in Tyl (CUU AGG C) has two unusual features. First, as noted above, it contains overlapping leucine codons in the 0 (CUU) and +1 (UUA) reading frames. Second, the arginine codon AGG within the frameshift signal is normally decoded by a tRNA, $tRNA^{Arg}_{CCU}$, which is present in low amounts in the host cell (Ikemura, 1982). Belcourt and Farabaugh (1990) demonstrated that increasing the intracellular levels of $tRNA^{Arg}_{CCU}$ in yeast decreases the level of frameshifting in Tyl, providing strong evidence that the absence of this tRNA from the ribosomal A-site enhances the frameshift event. These investigators also determined that slippage from the first to the overlapping, downstream leucine codon is essential to the frameshift event and presumably occurs with peptidyl-$tRNA^{Leu}$ (Belcourt and Farabaugh, 1990). They introduced a leucine tRNA gene with anticodon AAG into the host, whose gene product could decode CUU but not UUA codons. Transfer RNA^{Leu}_{AAG} would be expected to compete with the "shifty" tRNA, $tRNA^{Leu}_{UAG}$, which is capable of reading all six leucine codons (Weissenbach *et al.*, 1977). The level of frameshifting in Ty was severely inhibited by $tRNA^{Leu}_{AAG}$, demonstrating that slippage from CUU to UUA is essential for frameshifting in Ty. These observations led Belcourt and Farabaugh (1990) to propose a "peptidyl-tRNA slippage" model for frameshifting in the +1 direction in Ty. In this model, the relatively low abundance of $tRNA^{Arg}_{CCU}$ results in a pause in translation at an AGG codon. If this codon partially overlaps an upstream slippery CUUA sequence, then in the absence of an occupied A-site, the peptidyl-$tRNA^{Leu}_{UAG}$ (which is on the CUU codon in the P-site) slips one base foward to decode UUA. The slippage event establishes the +1 reading frame and normal translation then proceeds (Belcourt and Farabaugh, 1990).

As noted above, Ty3 does not have a sequence within a 21-bp region that promotes frameshifting comparable to the Tyl seven-nucleotide frameshift signal (S. Sandmeyer, personal communication 1991). Candidates for the frameshift signal within the 21-bp region involve an alanine GCG codon and an arginine CGA codon, which are used infrequently in yeast (and hence their cognate isoacceptors are present in low abundance) (see Ikemura, 1982). Either one of these codons may serve a similar function as AGG in the Tyl frameshift signal. The codon

immediately upstream of GCG is AAG (a lysine codon) and that of CGA is AAC (an asparagine codon), and a change to the +1 reading frame would mean that either $tRNA^{Lys}_{CUU}$ slips to misread GGA or $tRNA^{Asn}_{GUU}$ slips to misread ACC. With respect to the latter tRNA it is of interest to note that yeast tRNA lacks the highly modified Q base in the anticodon; as noted in Section III,A,5, $tRNA^{Asn}$ without Q base is proposed as the shifty isoacceptor in some of the −1 frameshift events. Ultimately, however, the site of the frameshift in Ty3 will have to be identified by mutagenesis studies and by sequencing the transframe protein.

IV. Concluding Remarks

The present review has focused on our present understanding of translational suppression in retroviral gene expression. The existence of this phenomenon obviously raises a number of important questions: What are the precise mechanisms of readthrough suppression and ribosomal frameshifting? What are the signals in the retroviral RNA that induce these unusual behaviors in the cellular translational machinery? Why have different retroviruses evolved completely distinct mechanisms that apparently accomplish the same ends? Finally, to what extent is either of these modes of translational suppression used as a regulatory mechanism in the synthesis of host cell proteins? As should be clear from the foregoing discussion, these questions in general remain to be answered; retroviruses appear to offer an invaluable tool for the analysis of translational mechanisms in higher eukaryotic cells.

Remarkably, there is currently no evidence that these suppression mechanisms are used by the host. It thus seems possible that an understanding of these phenomena will suggest approaches, including the use of antisense RNA or new types of antiviral drugs, which could help combat the induction of disease by these viruses.

Addendum

Since this review was prepared, mutational analysis of artificial constructs including the Mo-MuLV *gag–pol* junction has (1) shown that the 57 *pol* nucleotides immediately 3′ of the gag termination codon are necessary and sufficient for suppression, and (2) provided strong evidence that the two base-paired stems in the proposed pseudoknot struc-

ture in this region (ten Dam *et al.*, 1990) are crucial for suppression (Wills *et al.*, 1991; Feng *et al.*, 1992).

ACKNOWLEDGMENTS

The authors thank Ian Brierley, Mario Chamorro, David Garfinkel, Ray Gesteland, Anne-Lise Haenni, Lou Henderson, Alik Honigman, Stephen Inglis, Amos Panet, C. W. A. Pleij, Suzanne Sandmeyer, James Skuzeski, Edwin ten Dam, Harold Varmus, and Reed Wickner for communicating data prior to publication. We thank Cheri Rhoderick and Carol Shawver for preparation of the manuscript, and Richard Frederickson for assistance with graphics.

Research sponsored in part by the National Cancer Institute, DHHS, under contract No. NO1-CO-74101 with ABL. The contents of this publication do not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

Atkins, J. F., Weiss, R. B., and Gesteland, R. F. (1990). *Cell* **62,** 413–423.

Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R., and Gross, H. J. (1984a). *EMBO J.* **3,** 351–356.

Beier, H., Barciszewska, M., and Sickinger, H.-D. (1984b). *EMBO J.* **3,** 1091–1096.

Belcourt, M. F., and Farabaugh, P. J. (1990). *Cell* **62,** 339–352.

Bienz, M., and Kubli, E. (1981). *Nature (London)* **294,** 188–190.

Bienz, M., Kubli, E., Kohli, J., de Henau, S., Huez, G., Marbaix, G., and Grosjean, H. (1981). *Nucleic Acids Res.* **9,** 3835–3850.

Björk, G. R., Ericson, J. U., Gustafsson, C. E. D., Hagervall, T. G., Jönsson, Y. H., and Wikström, P. M. (1987). *Annu. Rev. Biochem.* **56,** 263–287.

Björk, G. R., Wikström, P. M., and Byström, A. S. (1989). *Science* **244,** 986–989.

Bossi, L. (1983). *J. Mol. Biol.* **164,** 73–87.

Bouzoubaa, S., Ziegler, V., Beck, D., Guilley, H., Richards, K., and Jonard, G. (1986). *J. Gen. Virol.* **67,** 1689–1700.

Brierley, I., Boursnell, M. E., Binns, M. M., Bilimoria, B., Blok, V. C., Brown, T. D. K., and Inglis, S. C. (1987). *EMBO J.* **6,** 3779–3785.

Brierley, I., Digard, P., and Inglis, S. C. (1989). *Cell* **57,** 537–547.

Brierley, I., Rolley, N. J., Jenner, A. J., and Inglis, S. C. (1991). *J. Mol. Biol.* **220,** 889–902.

Brown, C. M., Stockwell, P. A., Trotman, C. N. A., and Tate, W. P. (1990). *Nucleic Acids Res.* **18,** 6339–6345.

Buckingham, R. H., and Kurland, C. G. (1977). *Proc. Natl. Acad. Sci. U.S.A.* **74,** 5496–5498.

Capone, J. P., Sedivy, J. M., Sharp, P. A., and RajBhandary, U. L. (1986). *Mol. Cell. Biol.* **6,** 3059–3067.

Caskey, C. T., Beaudet, A., and Nirenberg, M. (1968). *J. Mol. Biol.* **37,** 99–118.

Cattaneo, R. (1989). *Trends Biochem. Sci.* **14,** 165–167.

Chakrabarti, L., Guyader, M., Alizon, M., Daniel, M. D., Desrosiers, R. C. Tiollais, P., and Sonigo, P. (1987). *Nature (London)* **328,** 543–547.

Chamorro, M., Parkin, N., and Varmus, H. E. (1992). *Proc. Natl. Acad. Sci. U.S.A.*, **89,**713–717.

Claesson, C., Samuelsson, T., Lustig, F., and Borén, T. (1990). *FEBS Lett.* **273,** 173–176.

Clare, J. J., and Farabaugh, P. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 2829–2833.
Clare, J. J., Belcourt, M., and Farabaugh, P. J. (1988). *Proc. Natl. Acad. Sci. U.S.A.* **85,** 6816–6820.
Coffin, J. M. (1990). *In* "Virology" (B. N. Fields, ed.), Vol. 2, pp. 1437–1500. Raven, New York.
Craigen, W. J., and Caskey, C. T. (1987). *Cell* **50,** 1–2.
Crawford, S., and Goff, S. P. (1985). *J. Virol.* **53,** 899–907.
Dayhuff, T. J., Atkins, J. F., and Gesteland, R. F. (1986). *J. Biol. Chem.* **261,** 7491–7500.
Dickson, C., Eisenman, R., Fan, H., Hunter, E., and Teich, N. (1984). *In* "Molecular Biology of Tumor Viruses" (R. Weiss, N. Teich, H. Varmus, and J. Coffin, eds.), 2nd Ed., pp. 513–648. Cold Spring Harbor Lab., Cold Spring Harbor, New York.
Dinman, J. D., Icho, T., and Wickner, R. B. (1991). *Proc. Natl. Acad. Sci. U.S.A.* **88,** 174–178.
Engelberg-Kulka, H. (1981). *Nucleic Acids Res.* **9,** 983–991.
Etzerodt, M., Mikkelsen, T., Pedersen, F. S., Kjeldgaard, N. O., and Jorgensen, P. (1984). *Virology* **134,** 196–207.
Feng, Y.-X., Hatfield, D. L., Rein, A., and Levin, J. G. (1989a). *J. Virol.* **63,** 2405–2410.
Feng, Y.-X., Levin, J. G., Hatfield, D. L., Schaefer, T. S., Gorelick, R. J., and Rein, A. (1989b). *J. Virol.* **63,** 2870–2873.
Feng, Y.-X., Copeland, T. D., Oroszlan, S., Rein, A., and Levin, J. G. (1990a). *Proc. Natl. Acad. Sci. U.S.A.* **87,** 8860–8863.
Feng, Y.-X., Levin, J. G., Hatfield, D. L., Schaefer, T. S., Gorelick, R. J., and Rein, A. (1990b). *In* "New Aspects of Positive-Strand RNA Viruses" (M. A. Brinton and F. X. Heinz, eds.), pp. 188–191. Am Soc. Microbiol., Washington, D.C.
Feng, Y.-X., Yuan, H., Rein, A., and Levin, J. G. (1992) Submitted.
Franchini, G., Gurgo, C., Guo, H. G., Gallo, R. C., Collalti, E., Fragnoli, K. A., Hall, L. F., Wong-Staal, F., and Reitz, Jr., M. S. (1987). *Nature (London)* **328,** 539–543.
French, B. T., Patrick, D. E., Grever, M. R., and Trewyn, R. W. (1991). *Proc. Natl. Acad. Sci. U.S.A.* **88,** 370–374.
Fujimura, T., and Wickner, R. B. (1988). *Cell* **55,** 663–671.
Garfinkel, D. J. (1991). *In* "The Retroviruses" (J. A. Levy, ed.). Plenum, New York. In press.
Geller, A. I., and Rich, A. (1980). *Nature (London)* **283,** 41–46.
Goelet, P., Lomonossoff, G. P., Butler, P. J. G. Akam, M. E., Gait, M. J., and Karn, J. (1982). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 5818–5822.
Graves, D. C., and Ferrer, J. F. (1976). *Cancer Res.* **36,** 4152–4159.
Guilley, H., Carrington, J. C., Balazs, E., Jonard, G., Richards, K., and Morris, T. J. (1985). *Nucleic Acids Res.* **13,** 6663–6677.
Guyader, M., Emerman, M., Sonigo, P., Clavel, F., Montagnier, L., and Alizon, M. (1987). *Nature (London)* **326,** 662–669.
Hansen, L. J., and Sandmeyer, S. B. (1990). *J. Virol.*. **64,** 2599–2607.
Hansen, L. J., Chalker, D. L., and Sandmeyer, S. B. (1988). *Mol. Cell. Biol.* **8,** 5245–5256.
Hatfield, D. (1972). *Proc. Natl. Acad. Sci. U.S.A.* **69,** 3014–3018.
Hatfield, D. (1985). *Trends Biochem. Sci.* **10,** 201–214.
Hatfield, D., and Oroszlan, S. (1990). *Trends Biochem. Sci.* **15,** 186–190.
Hatfield, D., Matthews, C. R., and Rice, M. (1979). *Biochim. Biophys. Acta* **564,** 414–423.
Hatfield, D., Diamond, A., and Dudock, B. (1982a). *Proc. Natl. Acad. Sci. U.S.A.* **79,** 6215–6219.
Hatfield, D., Varricchio, F., Rice, M., and Forget, B. G. (1982b). *J. Biol. Chem.* **257,** 3183–3188.

Hatfield, D., Feng, Y.-X., Lee, B. J., Rein, A., Levin, J. G., and Oroszlan, S. (1989). *Virology* **173,** 736–742.

Hatfield, D., Lee, B. J., Smith, D. W. E., and Oroszlan, S. (1990a). *In* "Progress in Molecular and Subcellular Biology" (P. Jeanteur, Y. Kuchino, W. E. G. Müller, and P. L. Paine, eds.), pp. 115–146. Springer-Verlag, New York.

Hatfield, D., Smith, D. W. E., Lee, B. J., Worland, P. J., and Oroszlan, S. (1990b). *Crit. Rev. Biochem. Mol. Biol.* **25,** 71–96.

Herr, W. (1984). *J. Virol.* **49,** 471–478.

Hiramatsu, K., Nishida, J., Naito, A., and Yoshikura, H. (1987). *J. Gen. Virol.* **68,** 213–218.

Hirsh, D. (1971). *J. Mol. Biol.* **58,** 439–458.

Hirsh, D., and Gold, L. (1971). *J. Mol. Biol.* **58,** 459–468.

Hizi, A., Henderson, L. E., Copeland, T. D., Sowder, R. C., Hixson, C. V., and Oroszlan, S. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84,** 7041–7045.

Honigman, A., Wolf, D., Yaish, S., Falk, H., and Panet, A. (1991), *Virology* **183,** 313–319.

Hughes, D., Thompson, S., O'Connor, M., Tuohy, T., Nichols, B. P., and Atkins, J. F. (1989). *J. Bacteriol* **171,** 1028–1034.

Icho, T., and Wickner, R. B. (1989). *J. Biol. Chem.* **264,** 6716–6723.

Ikemura, T. (1982). *J. Mol. Biol.* **158,** 573–597.

Inoue, J., Watanabe, T., Sato, M., Oda, A., Toyoshima, K., Yoshida, M., and Seiki, M. (1986). *Virology* **150,** 187–195.

Ishikawa, M., Meshi, T., Motoyoshi, F., Takamatsu, N., and Okada, Y. (1986). *Nucleic Acids Res.* **14,** 8291–8305.

Jacks, T. (1990). *Curr. Top. Microbiol. Immunol.* **157,** 93–124.

Jacks, T., Townsley, K., Varmus, H. E., and Majors, J. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84,** 4298–4302.

Jacks, T., Madhani, H. D., Masiarz, F. R., and Varmus, H. E. (1988a). *Cell* **55,** 447–458.

Jacks, T., and Varmus, H. E. (1985). *Science* **230,** 1237–1242.

Jacks, T., Power, M. D., Masiarz, F. R., Luciw, P. A., Barr, P. J., and Varmus, H. E. (1988b). *Nature (London)* **331,** 280–283.

Jackson, R. J., and Hunt, T. (1983). *Methods Enzymol.* **96,** 50–74.

Jamjoom, G. A., Naso, R. B., and Arlinghaus, R. B. (1977). *Virology* **78,** 11–34.

Jones, D. S., Nemoto, F., Kuchino, Y., Masuda, M., Yoshikura, H., and Nishimura, S. (1989). *Nucleic Acids Res.* **17,** 5933–5945.

Kato, S., Matsuo, K., Nishimura, N., Takahashi, N., and Takano, T. (1987). *Jpn. J. Genet.* **62,** 127–137.

Katoh, I., Yoshinaka, Y., Rein, A., Shibuya, M., Odaka, T., and Oroszlan, S. (1985). *Virology* **145,** 280–292.

Katze, J. R., Beck, W. T., Cheng, C. S., and McCloskey, J. A. (1983). *Recent Results Cancer Res.* **84,** 146–159.

Kawakami, T., Sherman, L., Dahlberg, J., Gazit, A., Yaniv, A., Tronick, S. R., and Aaronson, S. A. (1987). *Virology* **158,** 300–312.

Kawamoto, S., Yamamoto, S., Ueda, K., Nagahata, T., Chisaka, O., and Matsubara, K. (1990). *Biochem. Biophys. Res. Commun.* **171,** 1130–1136.

Kerr, I. M., Olshevsky, U., Lodish, H. F., and Baltimore, D. (1976). *J. Virol.* **18,** 627–635.

Kingsman, A. J., Gimlich, R. L., Clarke, L., Chinault, A. C., and Carbon, J. (1981). *J. Mol. Biol.* **145,** 619–632.

Kingsman, A. J., Wilson, W., and Kingsman, S. M. (1990). *In* "Retroviral Proteases" (L. H. Pearl, ed.), pp. 41–53. Stockton Press, New York.

Kohli, J., and Grosjean, H. (1981). *Mol. Gen. Genet.* **182,** 430–439.

Konecki, D. S., Aune, K. C., Tate, W., and Caskey, C. T. (1977). *J. Biol. Chem.* **252,** 4514–4520.

Kuchino, Y., Beier, H., Akita, N., and Nishimura, S. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84,** 2668–2672.

Le, S.-Y., Chen, J.-H., and Maizel, J. V. (1989). *Nucleic Acids Res.* **17,** 6143–6152.

Lee, B. J., Worland, P. J., Davis, J. N., Stadtman, T. C., and Hatfield, D. L. (1989). *J. Biol. Chem.* **264,** 9724–9727.

Lee, H.-J., Shieh, C.-K., Gorbalenya, A. E., Koonin, E. V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., and Lai, M. M. C. (1991). *Virology* **180,** 567–582.

Leis, J., Baltimore, D., Bishop, J. M., Coffin, J., Fleissner, E., Goff, S. P., Oroszlan, S., Robinson, H., Skalka, A. M., Temin, H. M., and Vogt, V. (1988). *J. Virol.* **62,** 1808–1809.

Levin, J. G., Hu, S. C., Rein, A., Messer, L. I., and Gerwin, B. I. (1984). *J. Virol.* **51,** 470–478.

Li, G. P., and Rice, C. M. (1989). *J. Virol.* **63,** 1326–1337.

Lin, J. P., Aker, M., Sitney, K. C., and Mortimer, R. K. (1986). *Gene* **49,** 383–388.

Lo, S. J., You, L.-R., and Lee, Y.-H. W. (1990). *Virology* **178,** 584–587.

Madhani, H. D., Jacks, T., and Varmus, H. E. (1988). *In* "Control of HIV Gene Expression" (R. Francz, B. Cullen, and F. Wong-Staal, eds.), pp. 119–125. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Mador, N., Panet, A., and Honigman, A. (1989). *J. Virol.* **63,** 2400–2404.

Marlor, R. L., Parkhurst, S. M., and Corces, V. G. (1986). *Mol. Cell. Biol.* **6,** 1129–1134.

Marshall, R. E., Caskey, C. T., and Nirenberg, M. (1967). *Science* **155,** 820–826.

Martin, R., Mogg, A. E., Heywood, L. A., Nitschke, L., and Burke, J. F. (1989). *Mol. Gen. Genet.* **217,** 411–418.

Meier, F., Suter, B., Grosjean, H., Keith, G., and Kubli, E. (1985). *EMBO J.* **4,** 823–827.

Mellor, J., Fulton, S. M., Dobson, M. J., Wilson, W., Kingsman, S. M., and Kingsman, A. J. (1985). *Nature (London)* **313,** 243–246.

Mietz, J. A., Grossman, Z., Lueders, K. K., and Kuff, E. L. (1987). *J. Virol.* **61,** 3020–3029.

Miller, A. D., and Verma, I. M. (1984). *J. Virol.* **49,** 214–222.

Miller, J. H., and Albertini, A. M. (1983). *J. Mol. Biol.* **164,** 59–71.

Miller, W. A., Waterhouse, P. M., and Gerlach, W. L. (1988). *Nucleic Acids Res.* **16,** 6097–6111.

Moore, R., Dixon, M., Smith, R., Peters, G., and Dickson, C. (1987). *J. Virol.* **61,** 480–490.

Murgola, E. J. (1990). *In* "Transfer RNAs and Other Soluble RNAs" (J. D. Cherayil, ed.), pp. 83–101. CRC Press, Boca Raton, Florida.

Murphy, E. C., Jr., Kopchick, J. J., Watson, K. F., and Arlinghaus, R. B. (1978). *Cell* **13,** 359–369.

Murphy, E. C., Jr., Wills, N., and Arlinghaus, R. B. (1980). *J. Virol* **34,** 464–473.

Nam, S. H., Kidokoro, M., Shida, H., and Hatanaka, M. (1988). *J. Virol.* **62,** 3718–3728.

Nam, S. H., Copeland, T. D., Hatanaka, M., and Oroszlan, S. (1992). Submitted.

Nirenberg, M., and Leder, P. (1964). *Science* **145,** 1399–1407.

Nishimura, S. (1979). *In* "Transfer RNA: Structure, Properties and Recognition" (P. Schimmel, D. Soll, and J. Abelson, eds.), Cold Spring Harbor Symp. Quant. Biol., Part A, pp. 59–79. Cold Spring Harbor Lab., Cold Spring Harbor, New York.

Panganiban, A. T. (1988). *J. Virol.* **62,** 3574–3580.

Pelham, H. R. B. (1978). *Nature (London)* **272,** 469–471.

Philipson, L., Andersson, P., Olshevsky, U., Weinberg, R., and Baltimore, D. (1978). *Cell* **13,** 189–199.

Pleij, C. W. A. (1990). *Trends Biochem. Sci.* **15,** 143–147.

Pleij, C. W. A., Rietveld, K., and Bosch, L. (1985). *Nucleic Acids Res.* **13,** 1717–1731.
Power, M. D., Marx, P. A., Bryant, M. L., Gardner, M. B., Barr, P. J., and Luciw, P. A. (1986). *Science* **231,** 1567–1572.
Pure, G. A., Robinson, G. W., Naumovski, L., and Friedberg, E. C. (1985). *J. Mol. Biol.* **183,** 31–42.
Raftery, L. A., Egan, J. B., Cline, S. W., and Yarus, M. (1984). *J. Bacteriol.* **158,** 849–859.
Randerath, E., Gupta, R. C., Chia, L. L. S. Y., Chang, S. H., and Randerath, K. (1979). *Eur. J. Biochem.* **93,** 79–94.
Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, Jr., S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, F. (1985). *Nature (London)* **313,** 277–284.
Rice, N. R., Stephens, R. M., Burny, A., and Gilden, R. V. (1985). *Virology* **142,** 357–377.
Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y., and Ikawa, Y. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 677–681.
Saigo, K., Kugimiya, W., Matsuo, Y., Inouye, S., Yoshioka, K., and Yuki, S. (1984). *Nature (London)* **312,** 659–661.
Sanchez-Pescador, R., Power, M. D., Barr, P. J., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., and Luciw, P. A. (1985). *Science* **227,** 484–492.
Sandmeyer, S. B., Hansen, L. J., and Chalker, D. L. (1990). *Annu. Rev. Genet.* **24,** 491–518.
Schimmel, P. (1989). *Cell* **58,** 9–12.
Schultze, M., Hohn, T., and Jiricny, J. (1990). *EMBO J.* **9,** 1177–1185.
Schwartz, D. E., Tizard, R., and Gilbert, W. (1983). *Cell* **32,** 853–869.
Sedivy, J. M., Capone, J. P., RajBhandary, U. L., and Sharp, P. A. (1987). *Cell* **50,** 379–389.
Seiki, M., Hattori, S., Hirayama, Y., and Yoshida, M. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 3618–3622.
Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S., Miwa, M., and Sugimura, T. (1985). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 3101–3105.
Shinnick, T. M., Lerner, R. A., and Sutcliffe, J. G. (1981). *Nature (London)* **293,** 543–548.
Skuzeski, J. M., Nichols, L. M., Gesteland, R. F., and Atkins, J. F. (1991). *J. Mol. Biol.* **218,** 365–373.
Smith, D. W. E., and Hatfield, D. L. (1986). *J. Mol. Biol.* **189,** 663–671.
Snijder, E. J., den Boon, J. A., Bredenbeek, P. J., Horzinek, M. C., Rijnbrand, R., and Spaan, W. J. M. (1990). *Nucleic Acids Res.* **18,** 4535–4542.
Sonigo, P., Alizon, M., Staskus, K., Klatzmann, D., Cole, S., Danos, O., Retzel, E., Tiollais, O., Haase, A., and Wain-Hobson, S. (1985). *Cell* **42,** 369–382.
Sonigo, P., Barker, C., Hunter, E., and Wain-Hobson, S. (1986). *Cell* **45,** 375–385.
Sprinzl, M., Hartmann, T., Weber, J., Blank, J., and Zeidler, R. (1989). *Nucleic Acids Res.* **17,** r1-r172.
Stephens, R. M., Casey, J. W., and Rice, N. R. (1986). *Science* **231,** 589–594.
Strauss, E. G., Rice, C. M., and Strauss, J. H. (1983). *Proc. Natl. Acad. Sci. U.S.A.* **80,** 5271–5275.
Strauss, E. G., Rice, C. M., and Strauss, J. H. (1984). *Virology* **133,** 92–110.
Strauss, E. G., Levinson, R., Rice, C. M., Dalrymple, J., and Strauss, J. H. (1988). *Virology* **164,** 265–274.
Tamura, T. (1983). *J. Virol.* **47,** 137–145.

ten Dam, E. B., Pleij, C. W. A., and Bosch, L. (1990). *Virus Genes* **4,** 121–136.
Thayer, R. M., Power, M. D., Bryant, M. L., Gardner, M. B., Barr, P. J., and Luciw, P. A. (1987). *Virology* **157,** 317–329.
Valle, R. P. C. (1989). Ph.D. Thesis, Univ. Paris VII.
Valle, R. P. C., and Haenni, A.-L. (1991). *In* "Translation in Eukaryotes" (H. Trachsel, ed.). pp. 177–189, CRC Press Inc., Boca Raton, FL.
Valle, R. P. C., and Morch, M.-D. (1988). *FEBS Lett.* **235,** 1–15.
Valle, R. P. C., Morch, M.-D., and Haenni, A.-L. (1987). *EMBO J.* **6,** 3049–3055.
Varmus, H., and Brown, P. (1989). *In* "Mobile DNA" (D. E. Berg and M. M. Howe, eds.), pp. 53–108. Am. Soc. Microbiol., Washington, D.C.
Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). *Cell* **40,** 9–17.
Weaver, T. A., Talbot, K. J., and Panganiban, A. T. (1990). *J. Virol.* **64,** 2642–2652.
Weiss, R., Lindsley, D., Falahee, B., and Gallant, J. (1988). *J. Mol. Biol.* **203,** 403–410.
Weiss, R. B., Dunn, D. M., Atkins, J. F., and Gesteland, R. F. (1987). *Cold Spring Harbor Symp. Quant. Biol.* **52,** 687–693.
Weiss, R. B., Dunn, D. M., Shuh, M., Atkins, J. F., and Gesteland, R. F. (1989). *New Biologist* **1,** 159–169.
Weiss, S. R., Hackett, P. B., Oppermann, H., Ullrich, A., Levintow, L., and Bishop, J. M. (1978). *Cell* **15,** 607–614.
Weiss, W. A., and Friedberg, E. C. (1986). *J. Mol. Biol.* **192,** 725–735.
Weiss, W. A., Edelman, I., Culbertson, M. R., and Friedberg, E. C. (1987). *Proc. Natl. Acad. Sci. U.S.A.* **84,** 8031–8034.
Weissenbach, J., Dirheimer, G., Falcoff, R., Sanceau, J., and Falcoff, E. (1977). *FEBS Lett.* **82,** 71–76.
Wickner, R. B. (1989). *FASEB J.* **3,** 2257–2265.
Wills, N. M., Gesteland, R. F., and Atkins, J. F. (1991). *Proc. Natl. Acad. Sci. U.S.A.* **86,** 409–413.
Wilson, R. K., and Roe, B. A. (1989). *Proc. Natl. Acad. Sci. U.S.A.* **86,** 409–413.
Wilson, W., Malim, M. H., Mellor, J., Kingsman, A. J., and Kingsman, S. M. (1986). *Nucleic Acids Res.* **14,** 7001–7016.
Wilson, W., Braddock, M., Adams, S. E., Rathjen, P. D., Kingsman, S. M., and Kingsman, A. J. (1988). *Cell* **55,** 1159–1169.
Wurch, T., Guidasci, T., Geldreich, A., Lebeurier, G., and Mesnard, J.-M. (1991). *Virology* **180,** 837–841.
Wyatt, J. R., Puglisi, J. D., and Tinoco, I., Jr. (1989). *BioEssays* **11,** 100–106.
Xiong, Z., and Lommel, S. A. (1989). *Virology* **171,** 543–554.
Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. (1985a). *Proc. Natl. Acad. Sci. U.S.A.* **82,** 1618–1622.
Yoshinaka, Y., Katoh, I., Copeland, T. D., and Oroszlan, S. (1985b). *J. Virol.* **55,** 870–873.