Proc. Natl. Acad. Sci. USA
Vol. 84, pp. 4298–4302, June 1987
Microbiology

Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus *gag*-related polyproteins

(protease gene/*pol*/translation/retrovirus/type D retrovirus)

TYLER JACKS*, KIM TOWNSLEY†, HAROLD E. VARMUS*, AND JOHN MAJORS†

*Departments of Biochemistry and Biophysics and Microbiology and Immunology, University of California School of Medicine, San Francisco, CA 94143; and †Department of Biochemistry, Washington University School of Medicine, St. Louis, MO 63110

Contributed by Harold E. Varmus, February 12, 1987

ABSTRACT **The primary translation products of retroviral *pol* genes are polyproteins initiated in an upstream gene (*gag*). To investigate the manner in which the *gag*-initiated polyproteins of the mouse mammary tumor virus are produced, we determined the nucleotide sequence of a 1.8-kilobase DNA fragment that spans the region between *gag* and *pol* in the C3H strain of mouse mammary tumor virus. The sequence reveals three overlapping open reading frames: the first encodes products of *gag* ($p27^{gag}$ and $p14^{gag}$); the second encodes a protein domain of unknown function (termed *X*) that is highly related to a similarly positioned sequence in simian type D retroviruses and the viral protease (*pro*); and the third encodes the reverse transcriptase. The reading frames are organized to permit uninterrupted readthrough from *gag* to *pol* if ribosomal frameshifts occur in the −1 direction within each of the two overlapping regions, one of which is 16 nucleotides in length and the other 13 nucleotides. Cell-free translation of RNA containing these overlap regions shows that fusion of the reading frames by ribosomal frameshifting occurs efficiently: about one-fourth of the ribosomes traversing the *gag–X/pro* overlap and one-tenth traversing the *X/pro–pol* overlap shift frames, generating *gag*-related polyproteins in ratios similar to those observed *in vivo*. Synthetic oligonucleotides containing either of the overlap regions inserted into novel contexts do not induce frameshifting; hence the overlapping portions of the reading frames are not sufficient to induce a frameshift event, and a larger sequence context or secondary structure may be implicated.**

The mouse mammary tumor virus (MMTV) is unusual among retroviruses in that it can propagate and act as a carcinogenic agent in mammary epithelial tissue, it is transcriptionally regulated by steroid hormones, and it has a type B morphology. Nevertheless it utilizes strategies for macromolecular synthesis similar to those observed with other retroviruses: (*i*) *gag*-encoded viral core proteins are coordinately synthesized as components of a large precursor protein that is subsequently processed by a virus-encoded protease; (*ii*) the *pol*-encoded reverse transcriptase and integrase proteins are expressed at lower levels by similar processing of a large, fused *gag–pol* precursor; and (*iii*) *env*-encoded glycoproteins are expressed from a spliced, subgenomic mRNA (for review, see ref. 1).

MMTV also has biochemical features that distinguish it from most other retroviruses: (*i*) the presence of a protein coding domain of unknown function in the long terminal repeat (2, 3); (*ii*) the assembly of core particles (type A particles) in the cytoplasm of infected cells (4); and (*iii*) the synthesis of three precursor polyproteins (Pr77, Pr110, and Pr160) that possess *gag* antigenic determinants (5–10). The latter two properties are shared with type D primate viruses, such as Mason–Pfizer monkey virus (11, 12).

The three MMTV *gag*-related polyproteins have been detected both in virus-infected cells and following cell-free translation of viral mRNAs and are present at ratios of ≈30:10:1 (5–7, 9, 10). Tryptic peptide and immunological analyses have shown that Pr77 is the *gag* precursor, which is processed by proteolytic cleavage to yield viral core proteins (p10, p21, p27, and p14; for review, see ref. 1). Pr110 is thought to result from COOH-terminal extension of Pr77 (5, 6). A minor core protein, p30, probably cleaved from Pr110, contains peptides derived both from the COOH terminus of Pr77 and from sequences unique to Pr110; therefore, the junction between Pr77 and Pr110 should lie within p30 (6, 8). By analogy with other retroviruses, Pr160 should result from extended synthesis into the MMTV *pol* region and should be processed into reverse transcriptase and integrase activities.

In several retroviral systems, synthesis of large *gag–pol* precursor proteins results from inefficient suppression (at the level of 3–5%) of translation termination signals at the end of the *gag* region, either by in-frame nonsense-codon suppression (13) or by translation frameshifting (14). The relative abundance of the three MMTV *gag* polyproteins suggests that, if termination suppression is used in their generation, it must be efficient. In the studies described here, we have determined the nucleotide (nt) sequence of an MMTV genome in the region of *gag* and *pol* and shown that it contains a third gene, *X/pro*, that lies between *gag* and *pol* and briefly overlaps them both. Using cell-free translation of SP6 RNA polymerase transcripts of the same DNA, we also show that the extended products are efficiently generated *in vitro* by ribosomal frameshifting.

MATERIALS AND METHODS

MMTV DNA. The substrate for sequence analysis was initially isolated as a 4-kilobase (kb) *Pst* I fragment from unintegrated circular MMTV DNA purified from rat XC cells infected with the C3H strain of MMTV (15). The fragment was cloned directly into the *Pst* I site of pBR322.

Sequence Analysis. Sequencing was done by the method of Maxam and Gilbert (16) using both sets of overlapping deletions generated by BAL-31 nuclease and subfragments generated with various restriction endonucleases. Sequence comparisons between MMTV and simian retrovirus-1 (SRV-1) were done using the program ALIGN. $p27^{gag}$, $p14^{gag}$, and *pro* domains were delineated from protein-sequencing data of Hizi and Oroszlan (personal communication).

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

Abbreviations: MMTV, mouse mammary tumor virus; SRV, simian retrovirus; RSV, Rous sarcoma virus; nt, nucleotide(s); HIV, human immunodeficiency virus; *Aha*-RNA, RNA generated by the restriction of pMGPP with *Aha* III; p10, p21, p27, and p14, viral core proteins of 10,000, 21,000, 27,000, and 14,000 daltons, respectively; Pr, precursor protein.

***In Vitro* Transcription and Translations.** SP6 transcription, rabbit reticulocyte translation, and immunoprecipitation reactions were done as described (14).

Plasmid Constructions. Plasmids were constructed as described in the figure legends using T4 DNA ligase (International Biotechnologies, New Haven, CT), *Escherichia coli* DNA polymerase I Klenow fragment (Boehringer Mannheim), and various restriction enzymes purchased from New England Biolabs. Oligonucleotides were synthesized by the Biomolecular Resource Center, University of California, San Francisco.

RESULTS AND DISCUSSION

The DNA and deduced protein sequences of the MMTV *gag–pol* region reveal three overlapping reading frames. To gain insight into the genetic organization of the MMTV *gag–pol* region, to assess the mechanism of synthesis of the nested polyproteins, and to study the relationship between MMTV and the type D retroviruses, we have determined the nucleotide sequence of a 1.8-kb DNA fragment that extends from midway through *gag* beyond the beginning of *pol*. The position of this *Pst* I–*Bgl* II fragment in the C3H MMTV genome is shown in Fig. 1.

Computer-assisted analysis of the nucleotide sequence reveals three extended and overlapping translational reading frames (Fig. 2). By comparing the amino acid sequences of these three reading frames with the sequences of other retroviral proteins and by taking into account the known pattern of MMTV polypeptide synthesis, we can assign each of the reading frames to known proteins. The first reading frame should code for the COOH terminus of $Pr77^{gag}$. By aligning our sequence with one previously deduced for the NH_2 terminus of the *gag* protein of the GR strain of MMTV (17), we were able to generate a hybrid GR/C3H sequence that predicts a *gag* protein with a molecular mass of 66 kDa, considerably smaller than its apparent molecular mass of 77 kDa in polyacrylamide gels. The portion of the *gag* sequence presented here codes for part of the major core protein $p27^{gag}$ and for all of $p14^{gag}$, a small nucleic acid-binding protein proteolytically cleaved from $Pr77^{gag}$ (18–20). Consistent with the postulated role for $p14^{gag}$, we find in its sequence two copies of the Cys-X_2-Cys-X_4-His-X_4-Cys peptide motif common to the small basic nucleic acid-binding proteins of all retroviruses (21).

Beginning 16 nucleotides upstream of the *gag* termination codon, in the −1 frame with respect to *gag*, is a second open reading frame that we have called *X/pro*. It extends for 304 codons beyond the *gag* terminus; fusion of the two frames would permit the synthesis of a protein of 95 kDa, again smaller than the experimentally determined size of Pr110. Contained within the COOH-terminal portion of this extended protein are two peptides, Asp-Thr-Gly-Ala-Asp and Gly-Arg-Asp, found in the presumed protease domains of most retroviruses (22). The protein domain, encoded by the first part of the second frame and called *X* in this discussion, is likely fused to $p14^{gag}$ to form the minor virion protein $p30^{gag-X}$ (5, 7).

A third reading frame begins 13 nucleotides upstream of the *X/pro* termination codon and should code for reverse transcriptase. Our sequence in this region is nearly identical to that previously determined for the beginning of the *pol* region of an MMTV endogenous element, *mtv-8* (23). The domain boundaries that we have drawn within our sequence are identical to those drawn by Moore *et al.* from a similar analysis of the sequence of the same region of the BR6 strain of MMTV (24).

FIG. 1. A physical and genetic map of the C3H MMTV genome. The nature of the *X/pro* domain is discussed in the text. Solid bar, *Pst* I–*Bgl* II fragment the sequence of which is presented; *orf*, open reading frame in the long terminal repeat (box).

Similarity of MMTV *gag*, *X*, and *pro* Sequences to Analogous Sequences in Type D SRVs. MMTV and the type D SRVs both make intracytoplasmic type A particles and also have similar patterns of *gag*-related proteins (4–12). To assess the sequence relatedness of proteins encoded by the *gag–X/pro* region of MMTV to those encoded by the equivalent region of a type D virus, we compared the amino acid sequences of the MMTV *gag* and *X/pro* reading frames with those of the corresponding region of the type D virus SRV-1 (25). [The homologies found with SRV-1 hold for another type D virus, Mason–Pfizer monkey virus, whose amino acid sequences are >97% identical to SRV-1 in the regions analyzed (26).] Fig. 3 (*Upper*) compares the amino acid sequences of the first open reading frame in our sequence. The major core protein domains, $p27^{gag}$ for MMTV and $p24^{gag}$ for SRV-1, show 33% amino acid identity. This contrasts with <20% identity between the NH_2-terminal *gag* protein domains of MMTV-GR and SRV-1 (data not shown). The nucleic acid-binding proteins ($p14^{gag}$) of the two viruses are more closely related (about 45% amino acid identity), with the second of the cysteine repeats being more similar than the first. Fig. 3 (*Lower*) compares the sequences of the second open reading frame, *X/pro*. The *pro* domains of the two viruses show >50% amino acid identity, as do the two *X* domains (with one identical peptide of 11 amino acids). The conserved nature of the *X* domains of the two viruses suggests that *X* has a similar and probably important role in their respective replication cycles, perhaps in the formation of intracytoplasmic A particles, an unusual property shared by these two viruses. Also, the homology to MMTV *X* begins in the *X/pro* reading frame of SRV-1 upstream of the *gag* termination codon. Thus, in the synthesis of the SRV-1 *gag–X/pro* fusion protein, Pr110, the transition from the *gag* to *X/pro* reading frame is likely to occur upstream of or within the last coding domain of *gag* (Fig. 3 *Lower*).

Frameshifting *in Vitro*. If Pr110 is the product of *gag* and *X/pro* and Pr160 is encoded by these genes plus *pol*, either mRNA splicing or ribosomal frameshifting must occur in order to align the reading frames. In Rous sarcoma virus (RSV), the protease is encoded mainly at the end of *gag*, and *pol* overlaps *gag* directly (27). We have recently used *in vitro* transcription and translation methods to show that frameshifting during translation occurs in the RSV *gag–pol* overlap to produce a *gag–pol* fusion protein at about 5% efficiency (14). We have used a similar strategy to demonstrate that the MMTV *X/pro* and *pol* genes are also expressed via translational frameshifting.

Fig. 4*A* shows relevant regions of the plasmid pMGPP, the template for *in vitro* transcription by SP6 RNA polymerase. An MMTV DNA fragment extending from the *Nde* I site near the end of *gag* to a *Hin*dIII site downstream of *pol* (Fig. 1) was inserted into an SP6 vector containing the 5′ portion of the RSV *gag* gene such that the two *gag* genes are in-frame. (The RSV *gag* segment provides the translational initiation codon and a convenient antigen for immunoprecipitation of the translation products.) Linearization of pMGPP at any of the restriction enzyme sites shown in Fig. 4*A*, followed by transcription with bacteriophage SP6 RNA polymerase, yields RNAs that extend to different points within the *X/pro* and *pol* genes. Restriction of pMGPP with *Aha* III generates an RNA (*Aha*-RNA) that includes the hybrid *gag* gene and, in the −1 frame, ≈1/3 of the *X/pro* domain. Normal translation of *Aha*-RNA will yield a 41-kDa *gag* protein.

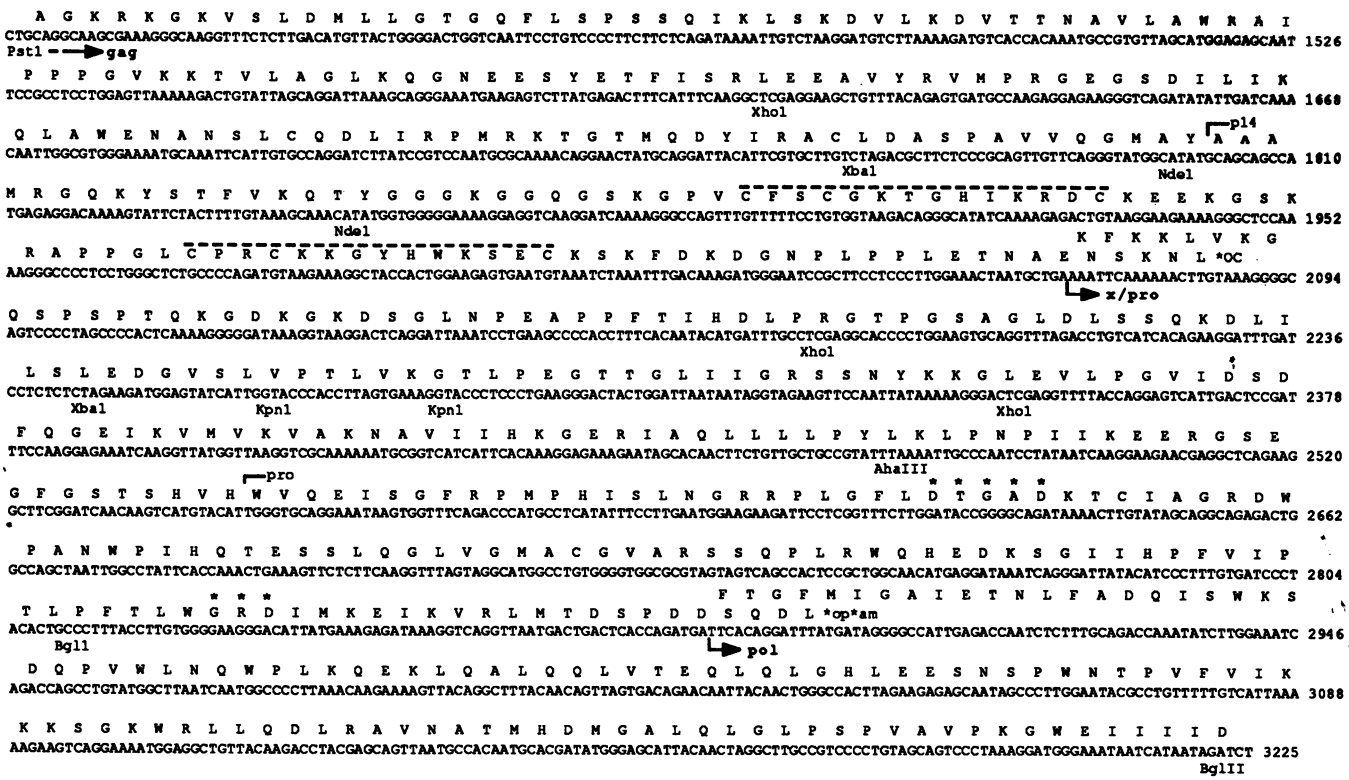

FIG. 2. DNA sequence of the *Pst* I–*Bgl* II fragment. Numbering is with respect to the start site of transcription as inferred from the sequence of Fasel *et al.* (17). The start points for the *X/pro* and *pol* open reading frames and the p14 and protease (pro) proteins as determined by Hizi and Oroszlan (personal communication) are shown, as are the cysteine-rich domains in p14 (----) and the conserved peptides in the protease domain (***). op, opal; am, amber; and oc, ochre termination codons.

However, if some fraction of translating ribosomes are able to shift into the −1 frame within the 16-nucleotide *gag–X/pro* overlap region, a 55-kDa *gag–X/pro* fusion protein will also be produced. Similarly, upon translation of the RNA synthesized from *Bgl* II-digested pMGPP (*Bgl*-RNA), a ribosomal frameshift within the *gag–X/pro* overlap will result in a full-length 70-kDa *gag–X/pro* fusion protein, and successive frameshifts, first at the *gag–X/pro* overlap and then at the *X/pro–pol* overlap (13 nucleotide and requiring a −1 shift), will yield a *gag–X/pro–pol* fusion protein of ≈82 kDa. The protein products of *Hin*d-RNA translation should be the same as those produced from *Bgl*-RNA, except that the *gag–X/pro–pol* fusion will be 120 kDa.

The size and distribution of the actual ^{35}S-labeled products of the translation of these RNAs in a rabbit reticulocyte lysate are exactly as predicted by efficient frameshifting at the *gag–X/pro* and *X/pro–pol* overlaps (Fig. 4*B*, lanes 1, 4, and 7). All of the proteins are precipitated by an anti-RSV p19^{gag} serum (lanes 2, 5, and 8), but not by nonimmune serum (lanes 3, 6, and 9).

We have estimated the efficiency of frameshifting at the *gag–X/pro* overlap by calculating the ratio of the *gag* to *gag–X/pro* proteins produced in the same translation. The amount of radioactivity in the excised gel slices, after correcting for differential methionine content, reveals a frameshifting efficiency of ≈23%. Of those ribosomes that do shift into the −1 frame at the *gag–X/pro* overlap, ≈8% also shift at the *X/pro–pol* overlap (data not shown). These efficiencies, while remarkably high, are consistent with those required to produce the observed levels of the *in vivo* analogues (6). Furthermore, Moore *et al.* (24) have employed a similar strategy using a DNA clone of MMTV-BR6 and also observe single- and double-frameshifting at efficiencies consistent with those reported here.

Three Types of Translational Suppression Control Synthesis of Retroviral *gag–pol* Proteins. Our demonstration that the MMTV *gag*-fusion proteins are produced via single- and double-frameshifting events provides the third type of translational control over the synthesis of retroviral *gag* fusion proteins. A single ribosomal frameshift is sufficient to express the coding potential of the RSV (14, 27) and human immunodeficiency virus (HIV) *gag–pol* regions (28–30; T.J. and H.E.V., unpublished work). Yoshinaka *et al.* (13, 31) have shown that suppression of an amber codon separating the *gag* and *pol* domains of murine leukemic virus (MuLV) and feline leukemia virus (FeLV) occurs to synthesize a *gag–pol* fusion protein encoded in a single reading frame.

The existence of a separate open reading frame between *gag* and *pol* to encode the viral protease is not unique to MMTV. Bovine leukemia virus (BLV), human T-cell leukemia virus type 2 (HTLV-2), and the type D simian viruses (Mason–Pfizer monkey virus, SRV-1, and SRV-2) also have a three-tiered *gag–pro–pol* arrangement (25, 26, 32, 33). Experimental verification of the predicted frameshifts, however, is not yet available.

Determining the Signals for Frameshifting: The Overlap Nucleotides Are Not Sufficient. We presume that the highly efficient frameshifting observed in the MMTV overlaps occurs in response to one or more special codons that are either themselves unusual or are in unusual contexts. Moreover, "frameshift signals" might exist in the overlapping reading frames of other viruses that have been shown to, or are believed to, utilize frameshifting. In fact, a simple nucleotide sequence search has uncovered two potential signals. The MMTV *gag–X/pro* overlap includes the sequence A AAA AAC (where the triplets denote *gag* codons) (Fig. 5*A* and *C*); this sequence is also present in the upstream overlaps of BLV (32) and HTLV-2 (33). Furthermore, the amino acid sequence of the MMTV protein p30 as determined

```
                                  [------p27
SRVGAG    GQHIPKGKCCASREKEEQTPKDIFPVT---ETVDGQGQAVRHHNGFDFTVIKELKTAASQ
MTVGAG    ------GAMAEAREKGDLTFT--FPVVFMGESDDDDTPVW---EPLPLKTLKELQSAVRT
Common          G     REK   T    FPV    E  D     W           KEL  A

SRVGAG    YGATAPYTLAIVESVADNWLTPTDWNTLVRAVLSGGDHLLWKSEFFENCRETAKR--NQQ
MTVGAG    MGPSAPYTLEVVDMVASQWLTPSDWHQTARATLSRGDYVLWRSEYEEKSKETVQKPAGKR
Common     G  APYTL  V  VA  WLTP DW    RA LS GD  LW SE  E   ET

SRVGAG    AGNGWDFDMLTGSGNYSSTDAQMQYDPGLFAQIQAAATKAWRKLPVKADPGASLTGVKQG
MTVGAG    KGKV-SLDMLLGTGQFLSPSSQIKLSKDVLKDVTTNAVLAWRAIPPPGVKKTVLAGLKQG
Common     G     DML G G   S   Q              A  AWR  P        L G KQG

SRVGAG    PDDGFADFVHRLITTAGRIFGSAEAGVDYVKQLAYENANPACQAAIRPYRKKTDLTGYIR
MTVGAG    NEESYETFISRLEETVYRVMPRGEGSDILIKQLAWENANSLCQDLIRPMRKTGTIQDYIR
Common          F  RL  T  R     E      KQLA ENAN  CQ  IRP RK       YIR

                          [--------p14
SRVGAG    LCSDIGPSYQQGLAMAAAFSGQTVKDFLNNK---NKE----KGGCCFKCGRKGHFAKNCH
MTVGAG    ACLDASPAVVQGMAYAAAMRGQKYSTFVKQTYGGGKGGQGSKGPVCFSCGKTGHIKRDCK
Common     C D  P   QG A AAA  GQ    F        K     KG  CF CG  GH    C
                                                      *--*----*----*

SRVGAG    EHIHNNSETKAPGLCPRCKRGKHWANECKSKTDSQGNPLPPHQGNGLRGQPQAPKQAYGA
MTVGAG    EE--KGSKRAPPGLCPRCKKGYHWKSECKSKFDKDGNPLPP-----LETNAENSKN----
Common    E     S     PGLCPRCK G HW  ECKSK D  GNPLPP     L       K
                      *--*----*----*

SRVGAG    VSFVPANKNNPFQSLPEPPQEVQDWTSVPPPTQY
MTVGAG    --------------L--------------------
Common                  L

          [--------x
SRVXPRO   --RKPTTTPSGKRTEGPAPGPETSLWGGQLCSSQQKQP--ISKLTRATPGSAGLDLSSTS
MTVXPRO   KFKKLVKGQSPSPTQKGDKGKDSGL-------NPEAPPFTIHDLPRGTPGSAGLDLSSQK
Common      K     S   T    G    L           P  I  L R TPGSAGLDLSS

          *
SRVXPRO   HTVLTPEMGPQALSTGIYGPLPPNTFGLILGRSSITIKGLQVYPGVIDNHYTGEIKIMAK
MTVXPRO   DLILSLEDGVSLVPTLVKGTLPEGTTGLIIGRSSNYKKGLEVLPGVIDSDFQGEIKVMVK
Common      L  E G     T   G LP  T GLI GRSS   KGL V PGVID    GEIK M K

                                                [----pro
SRVXPRO   AVNNIVTVPQGNRIAQLILLPLIETDNKVQQPYRGQGSFGS-SDIYWVQPITCQKPSLTL
MTVXPRO   VAKNAVIIHKGERIAQLLLLPYLKLPNPIIKEERGSEGFGSTSHVHVVQEISGFRPMPHI
Common      N V    G RIAQL LLP     N      RG   FGS S   WVQ I    P

SRVXPRO   WLDDKMFTGLIDTGADVTIIKLEDVPPNWPITDTLTNLRG-IGQSNNPKQSSKYLTWRDK
MTVXPRO   SLNGRRFLGFLDTGADKTCIAGRDVPANWPIHQTESSLQGLVGMACGVARSSQPLRWQH-
Common     L    F G  DTGAD T I   DWP NWPI  T   L G  G       SS  L W
                    #####

SRVXPRO   ENNSGLIKPFVIPNLPVNLWGRDLLSQMKIMMCSPSDIVTAQMLAQGYSPGKGLGKNENG
MTVXPRO   EDKSGIIHPFVIPTLPFTLWGRDIMKEIKV-----------RLMTDS-------------
Common    E  SG I PFVI  LP  LWGRD     K
                        ###

SRVXPRO   ILHPIPNQGQFDKKGFGNF
MTVXPRO   -----PDDSQ-DL------
Common         P   Q D
```

FIG. 3. (*Upper*) A comparison of the $p27^{gag}$ and $p14^{gag}$ domains of MMTV (MTVGAG) and a type D virus, SRV-1 (SRVGAG) (25). The MMTV p27 protein sequence is a hybrid GR/C3H sequence derived by joining the sequence of Fasel *et al.* (17) with ours. The NH_2 termini of MMTV $p27^{gag}$ and $p14^{gag}$ are shown, and the repeated cysteine-rich motif in p14 is indicated (*----*). (*Lower*) A comparison of the amino acid sequences of the *X/pro* domains of MMTV (MTVXPRO) and SRV-1 (SRVXPRO). The asterisk (*) above the valine in the SRV sequence indicates the end of the SRV-1 *gag* reading frame (25). The conserved *X* sequence TPGSAGLDLSS lies upstream of this site. Conserved *pro* peptides are indicated by (#).

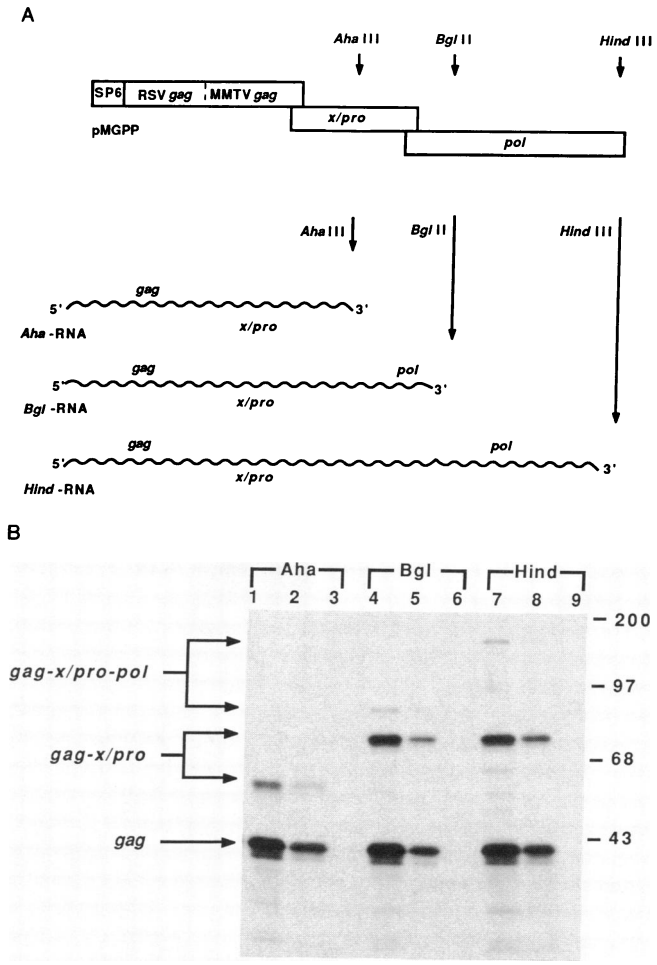

FIG. 4. (*A*) The plasmid pMGPP was constructed by ligating a 3.2-kb DNA fragment extending from a *Nde* I site within the p27 domain of MMTV *gag* (position 1796 in Fig. 2; also see Fig. 1) to a *Hin*dIII site downstream of *pol* (Fig. 1) to the plasmid pGP (14), previously digested with *Nde* I and *Hin*dIII. pGP is an SP6 vector containing the *gag–pol* domain of RSV; the *Nde* I site is in the p27 domain of *gag* (position 1290 in ref. 27), and the *Hin*dIII site is in the downstream polylinker. Cleavage of pMGPP with *Aha* III, *Bgl* II, or *Hin*dIII, followed by *in vitro* transcription with SP6 polymerase, generates the three mRNA species shown. (*B*) Fluorogram of ^{35}S-labeled proteins produced from rabbit reticulocyte lysate translation of *Aha*-, *Bgl*-, and *Hin*d-RNAs. (Lanes 1, 4, and 7) Unprecipitated proteins; (lanes 2, 5, and 8) proteins precipitated with anti-RSV $p19^{gag}$ serum; (lanes 3, 6, and 9) proteins precipitated with nonimmune rabbit serum. The positions of the expected products are indicated by arrows, and the positions of the molecular mass markers are indicated in kDa.

by Hizi and Oroszlan is consistent with the *gag–X/pro* frameshift occurring at this sequence (personal communication). The second potential signal, U UUA, is present in the downstream overlaps of MMTV (Fig. 5*A* and *C*), BLV (32), and HTLV-2 (33), as well as the single overlaps of RSV (27) and HIV (28–30). We have recently used amino acid sequencing and site-directed mutagenesis to show that this sequence is the frameshift site in RSV RNA (T.J., F. Masiarz, H.E.V., unpublished work). Given these potential signals and the fact that all of the retroviral frameshifts are in the −1 direction, a simple model for frameshifting would call for the tRNA reading the 0-frame codon (AAA, AAC, or UUA) to slip back one nucleotide and pair with the codon in the −1 frame.

Regardless of the details of the frameshifting mechanism, it seemed possible that all of the information required for frameshifting would reside in the signals described above or in the signals plus the adjacent nucleotides within the overlapping portions of the reading frames. To test whether the overlaps are sufficient to induce frameshifting, we cloned synthetic oligonucleotides corresponding to the two MMTV overlaps in between two new genes. As shown in Fig. 5*A*, the *gag–X/pro* and *X/pro–pol* overlaps join a portion of the 5′ end of the RSV *gag* gene and part of the 3′ region of the HIV *pol* gene in the plasmids pOL1 and pOL2; the plasmids are constructed so that the production of a *gag–pol* fusion protein is dependent on −1 frameshifting in the overlap segments.

The fluorogram in Fig. 5*B* shows the unprecipitated products of the translation of pOL1 and pOL2 RNAs. Despite copious amounts of the expected 43-kDa *gag* protein, neither RNA yields significant amounts of the *gag–pol* fusion (predicted to be 53 kDa). It appears, therefore, that in this new context the MMTV overlaps are not sufficient to promote efficient frameshifting. The simplest explanation for this failure is that sequences that border the overlaps in wild-type MMTV mRNA are involved in the frameshifting process.

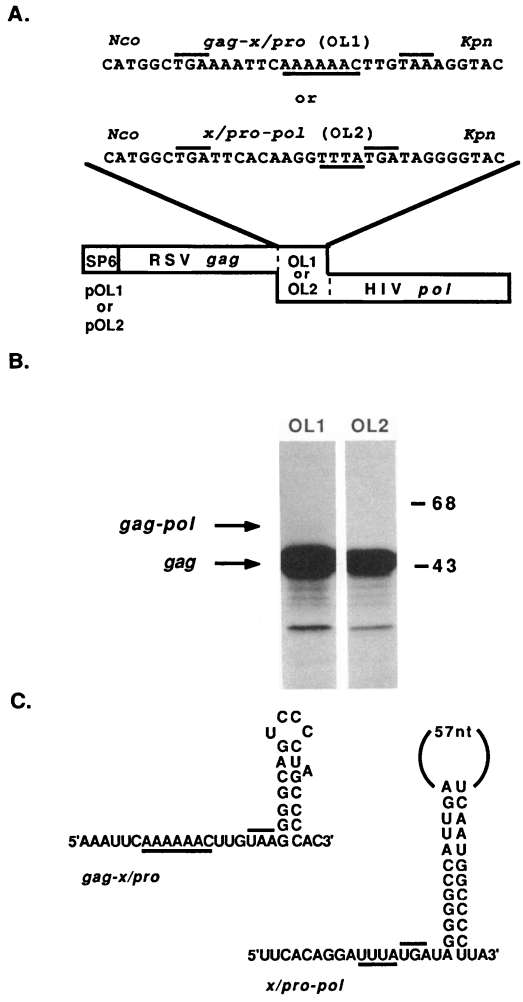

FIG. 5. (*A*) The plasmids pOL1 and pOL2. Synthetic oligonucleotides containing the plus strand of the *gag–X/pro* (pOL1) or *X/pro–pol* (pOL2) overlaps were ligated between an *Nco* I site in the 5′ portion of the RSV *gag* and a *Kpn* I site in the 3′ portion of the HIV *pol* in an SP6 vector. (The oligonucleotides were synthesized with *Nco* I- and *Kpn* I-compatible ends.) Ligation was followed by filling in the single-stranded region using the Klenow fragment of *E. coli* DNA polymerase I. The DNA sequences in the region of the overlaps were verified using the method of Chen and Seeburg (34). The termination codons that delineate the overlaps are overlined, and the proposed frameshift signals are underlined. (*B*) Fluorogram of unprecipitated, ^{35}S-labeled proteins produced from rabbit reticulocyte lysate translation of pOL1 and pOL2 RNAs. The predicted positions of the *gag* and *gag–pol* proteins are shown (arrows), and the positions of molecular mass markers are indicated in kDa. (*C*) Potential stem-loop structures located 3′ to the MMTV *gag–X/pro* and *X/pro–pol* overlaps. The complete overlaps are shown, with the *gag* and *X/pro* termination codons overlined and the proposed frameshift signals underlined.

Alternatively, there could be a negative effect on frameshifting exerted by the sequences that now surround the overlaps. Interestingly, just 3′ to both MMTV overlaps in their natural setting are potential stem-loop structures. Should these stem-loop structures be involved in frameshifting, they may act by stalling translating ribosomes, thereby promoting the tRNA slippage postulated above.

We thank Clive Dickson, Paul Luciw, Gordon Peters, and Steve Oroszlan for communicating results prior to publication and Janine Marinos for help with preparation of the manuscript. This work was supported by a University of California, San Francisco Earle C. Anthony/Dean's Fellowship (T.J.) and Public Health Service Grants CA38994 (J.M.) and CA39832 (H.E.V.). H.E.V. is an American Cancer Society Research Professor.

1. Weiss, R., Teich, N., Varmus, H. & Coffin, J., eds. (1982) *RNA Tumor Viruses* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
2. Dickson, C., Smith, R. & Peters, G. (1981) *Nature (London)* **291,** 511–513.
3. Donehower, L. A., Huang, A. L. & Hager, G. L. (1981) *J. Virol.* **37,** 226–238.
4. Bernhard, W. (1958) *Cancer Res.* **18,** 491–509.
5. Anderson, S. J., Naso, R. B., Davis, J. & Bowen, J. M. (1979) *J. Virol.* **32,** 507–516.
6. Dickson, C. & Atterwill, M. (1979) *Cell* **17,** 1003–1012.
7. Dickson, C. & Peters, G. (1981) *J. Virol.* **37,** 36–47.
8. Gautsch, J., Lerner, R., Howard, D., Teramoto, Y. & Schlom, J. (1978) *J. Virol.* **27,** 688–699.
9. Massey, R. J. & Schochetman, G. (1979) *Virology* **99,** 358–371.
10. Sen, G. C., Racevskis, J. & Sarkar, N. G. (1981) *J. Virol.* **37,** 963–975.
11. Bradac, J. & Hunter, E. (1984) *Virology* **138,** 260–275.
12. Chatterjee, S., Bradac, J. & Hunter, E. (1985) *Virology* **141,** 65–76.
13. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 1618–1622.
14. Jacks, T. & Varmus, H. E. (1985) *Science* **230,** 1237–1242.
15. Majors, J. & Varmus, H. E. (1981) *Nature (London)* **289,** 253–258.
16. Maxam, A. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 560–564.
17. Fasel, N., Buetti, E., Firzlaff, J., Pearson, K. & Diggelmann, H. (1983) *Nucleic Acids Res.* **11,** 6943–6955.
18. Arthur, L., Long, C., Smith, G. & Fine, D. (1978) *Int. J. Cancer* **22,** 433–440.
19. Dickson, C. & Peters, G. (1983) *Curr. Topics Microbiol. Immunol.* **106,** 1–34.
20. Long, C. W., Berzinksi, R. & Gilden, R. V. (1977) *Int. J. Cancer* **19,** 843–850.
21. Copeland, T., Oroszlan, S., Kalyanaraman, V., Sarngadharan, M. & Gallo, R. (1983) *FEBS Lett.* **162,** 390–395.
22. Toh, H., Ono, M., Saigo, K. & Miyata, T. (1985) *Nature (London)* **315,** 691.
23. Deen, K. & Sweet, R. (1986) *J. Virol.* **57,** 422–432.
24. Moore, R., Dixon, M., Smith, R., Peters, G. & Dickson, C. (1987) *J. Virol.* **61,** 480–490.
25. Power, M., Marx, P., Bryant, M., Gardner, M., Barr, P. & Luciw, P. (1986) *Science* **231,** 1567–1572.
26. Sonigo, P., Barker, C., Hunter, E. & Wain-Hobson, S. (1986) *Cell* **45,** 375–385.
27. Schwartz, D. E., Tizard, R. & Gilbert, W. (1983) *Cell* **32,** 853–869.
28. Ratner, L., Haseltine, W., Patarca, R., Livzk, K. J., Starcich, R., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Jr., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C. & Wong-Staal, F. (1985) *Nature (London)* **313,** 277–284.
29. Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S. & Alizon, M. (1985) *Cell* **40,** 9–17.
30. Sanchez-Pescador, R., Power, M. D., Barr, P., Steimer, K. S., Stempien, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D. & Luciw, P. A. (1985) *Science* **227,** 484–492.
31. Yoshinaka, Y., Katoh, I., Copeland, T. D. & Oroszlan, S. (1985) *J. Virol.* **55,** 870–873.
32. Rice, N. R., Stephens, R. M., Burny, A. & Gilden, R. V. (1985) *Virology* **142,** 357–377.
33. Shimotohno, K., Takahashi, Y., Shimizu, N., Gojobori, T., Golde, D. W., Chen, I. S. Y., Miwa, M. & Sugimura, T. (1985) *Proc. Natl. Acad. Sci. USA* **82,** 3101–3105.
34. Chen, E. Y. & Seeburg, P. H. (1985) *DNA* **4,** 165–170.