JOURNAL OF VIROLOGY, May 1989, p. 2400–2404
0022-538X/89/052400-05$02.00/0

Vol. 63, No. 5

Translation of *gag*, *pro*, and *pol* Gene Products of Human T-Cell Leukemia Virus Type 2

NURITH MADOR, AMOS PANET, AND ALIK HONIGMAN*

Departments of Molecular Genetics and Virology, The Hebrew University-Hadassah Medical School, Jerusalem, 91010 Israel

Received 28 October 1988/Accepted 2 February 1989

Sequence analysis of human T-cell leukemia proviral DNA revealed three open reading frames arranged at a −1 position relative to one another. On the basis of homology to other retroviruses, these open reading frames were assigned to the *gag*, *pro*, and *pol* genes. To characterize the primary protein products of these genes and their modes of synthesis, a DNA clone of human T-cell leukemia virus type 2 was transcribed and translated in vitro. Analysis of the viral proteins revealed three polyproteins with molecular masses of 58, 75, and 112 kilodaltons at relative frequencies of 100:13:0.9, respectively. These proteins were mapped on the viral genome by both internal deletions and 3′-end truncations at *gag*, *pro*, and *pol*, respectively. The results indicate that translation of the *pol* gene requires two independent frameshift events, and the readthrough frequencies at the two frameshift sites appeared to be similar.

Sequence analysis of the proviral DNA of human T-cell leukemia viruses (HTLV) has indicated a unique arrangement of the *gag* and *pol* genes. While in most groups of retroviruses these two adjacent genes are separated by only a translation termination codon, an open reading frame (ORF) with a coding capacity for a protein of either 22 or 29 kilodaltons (kDa) is found between the *gag* and *pol* genes of HTLV types 2 and 1 (HTLV-2 and HTLV-1), respectively (2, 8, 13, 16, 18). On the basis of sequence relatedness among retroviruses, this ORF appears to code for the viral protease and is therefore termed the *pro* gene. Expression of the protease as part of the *pol* gene is modulated at the level of translation by either of the following two mechanisms: (i) in murine leukemia viruses by infrequent insertion of a glutamine at an amber UAG codon present between the *gag* and *pol* genes (9, 20) or (ii) through a ribosomal frameshift at a defined site to read through two ORFs which are arranged at −1 with respect to one another (4, 5, 7). Both translation control mechanisms yield the Gag-Pol polyprotein and Gag at ratios ranging from about 1:8 (human immunodeficiency virus) to 1:20 (Rous sarcoma virus and murine leukemia virus) (5, 7, 12). HTLV, bovine leukemia virus, and mouse mammary tumor virus (MMTV) may represent yet a third group of viruses with respect to expression of the *pro* and *pol* genes (2, 3, 6, 11, 14, 18). Translation in vitro of MMTV RNA, in fact, yielded three polyproteins which may represent Gag, Gag-Pro, and Gag-Pro-Pol. Expression of the *pol* gene in MMTV appears to require two successive frameshift events. The first shift in the reading frame generates the Gag-Pro polyprotein, and the second leads to the synthesis of Gag-Pro-Pol (3, 6, 11).

There is little experimental data with regard to the synthesis of HTLV proteins. In chronically HTLV-1-infected cells, the 53-kDa *gag* polyprotein was detected but no Gag-Pro or Gag-Pro-Pol products were observed (1).

To identify the HTLV-2 gene products, a proviral DNA which compromises the putative ORFs of *gag* and *pro* and part of that of *pol* (nucleotides 361 to 3633; 17, 18) was inserted at the polylinker of pSP65 (10), giving rise to plasmid pNM10 (Fig. 1). The plasmid was cleaved with endonuclease *Hin*dIII at the polylinker, downstream from the inserted HTLV-2 DNA, and served as a template in the SP6 RNA polymerase transcription system (10). The resulting uniform-size 3.2-kilobase RNA (HIND-RNA) was added to an in vitro rabbit reticulocyte translation system supplemented with [^{35}S]methionine. Samples were withdrawn at intervals, and polypeptides were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (Fig. 2). A major product with an apparent molecular mass of 58 kDa was observed after only 10 min of translation. The calculated molecular mass of the protein encoded by the first ORF was 55 kDa. The 75-kDa product observed after 20 and 40 min of translation may correspond to a fused protein with a calculated molecular mass of 76 kDa encoded by joining the first two ORFs of HTLV-2. Since the HTLV-2 DNA was cleaved at the *Xba*I restriction site (nucleotide 3633; Fig. 1), only part of the third ORF is included in HIND-RNA. The calculated size of the polyprotein encoded by the three ORFs in pNM10 is 120 kDa. The minor polypeptide with an apparent size of 112 kDa observed after 20 and 40 min of translation corresponds in size to the three joined ORFs. In addition to the translation products corresponding in size to proteins expected from the HTLV-2 DNA sequence, other polypeptides were observed (Fig. 2). The origin of these bands is discussed below.

To map the three proteins described above on the HTLV-2 genome, the DNA was cleaved within each of the three putative *gag*, *pro*, and *pol* genes before in vitro transcription (Fig. 3). Restriction endonuclease *Nco*I cleaved within the putative *gag* gene to give rise to a 1.1-kilobase transcript (NCO-RNA). Transcription of pNM10 after cleavage with endonuclease *Nar*I gave rise to 1.9-kilobase RNA (NAR-RNA) with the 3′ end located within the putative *pro* ORF. An RNA of 2.4 kilobases was obtained following transcription of pNM10 cleaved with endonuclease *Nde*I within the putative *pol* gene (NDE-RNA). Homogeneous RNA species were obtained with a common 5′ end initiated at the SP6 promoter and 3′ ends terminated at the corresponding restriction enzyme cleavage sites of the template DNA (Fig. 3A). If the three polypeptides described above initiate at the same AUG codon, translation of the truncated RNAs is expected to yield shorter polypeptides (Fig. 3B). Translation

* Corresponding author.

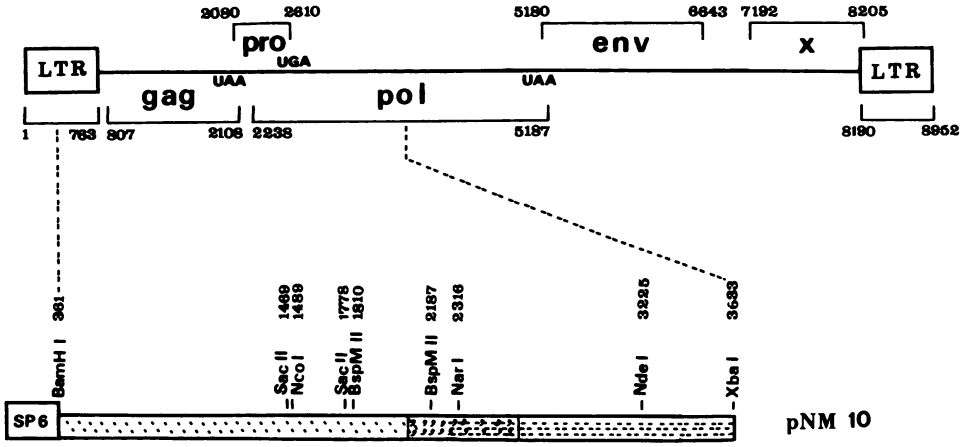

FIG. 1. Physical map of HTLV-2 provirus. Numbers indicate nucleotides from the beginning of the 5′ long terminal repeat (LTR) (18). The translation termination codons of *gag*, *pro*, and *pol* are indicated as on the mRNA. The HTLV-2 DNA fragment bordered by the *Bam*HI (nucleotide 361) and *Xba*I (nucleotide 3633) restriction sites was isolated from clone λMOISA (17) and inserted into pSP65 (bottom line). The endonucleases used are indicated above their restriction sites.

of NCO-RNA should yield a 30-kDa protein. In fact, a protein with an apparent mass of 35 kDa replaced the 58-kDa Gag protein (compare lanes 1 and 4). This novel product migrated more slowly than the 34-kDa protein, which appeared to be initiated at an internal AUG codon within the *gag* gene (compare lanes 1 and 1′ with lanes 2, 3, and 4). The discrepancy between the apparent 35-kDa polypeptide and the calculated 30-kDa polypeptide may be due to nonlinear migration of high- and low-molecular-weight proteins.

Translation of NAR-RNA is expected to yield a 58-kDa intact Gag protein and a truncated Gag-Pro protein of 64 kDa. The 58- and 68-kDa proteins (Fig. 3B, lane 2) appear to correspond in size to the two expected proteins.

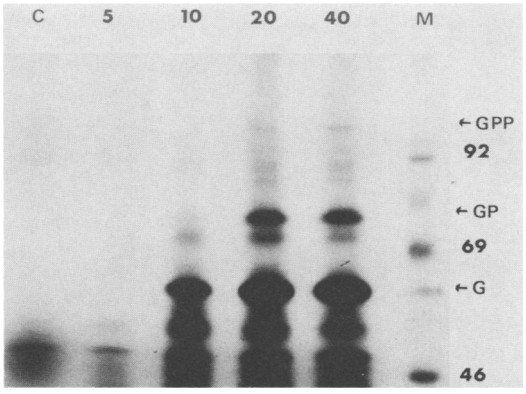

FIG. 2. Translation kinetics of RNA transcribed from HTLV-2 proviral DNA. Plasmid pNM10 was cleaved in the polylinker with endonuclease *Hin*dIII and transcribed by SP6 RNA polymerase (New England BioLabs, Inc.) in the presence of ribonucleotide triphosphates (Promega Biotec) (10). HIND-RNA (2 ng) was denaturated for 2 min at 68°C and translated in 25 μl of nuclease-treated rabbit reticulocyte lysate (Promega Biotec) supplemented with [^{35}S]methionine (50 μCi; Amersham Corp). Following translation, samples (10 μl each) were withdrawn after 5, 10, 20, and 40 min for analysis on a sodium dodecyl sulfate–10% polyacrylamide gel. Lanes: C, translation (40 min) without added exogenous mRNA; M, molecular mass markers (sizes are indicated in kilodaltons on the right). The putative gene products Gag (G), Gag-Pro (GP), and Gag-Pro-Pol (GPP) are indicated. Scanning of autoradiograms was performed on several time exposures of each gel autoradiogram by using a computerized photoscanner (620 Video Densitometer; Bio-Rad Laboratories).

Translation through the three ORFs in NDE-RNA yielded a truncated Gag-Pro-Pol polyprotein of 102 kDa (lane 3) which replaced the 112-kDa protein product of HIND-RNA (Fig. 3B, lane 4). The barely visible 112-kDa band (lane 4, arrow) is identical to the protein shown in Fig. 2 (lanes 20 and 40, GPP). In addition, a full-length 75-kDa Gag-Pro fusion protein and a 58-kDa Gag polypeptide were evident after translation of NDE-RNA (Fig. 3B, lane 4).

Thus, translation of RNAs truncated within each of the three ORFs enabled us to correlate the translation products with the HTLV-2 *gag*, *pro*, and *pol* genes.

Several additional translation products were seen in the autoradiograms presented in Fig. 2 and 3B. Two of these products are endogenous reticulocyte proteins on the basis of the translation reactions without addition of exogenous RNA (Fig. 2 and 3B, lanes C). Several reasons may account for other relatively short polypeptides observed in the autoradiograms: (i) initiation at internal AUG codons, (ii) early quitting of nascent polypeptide chains, and (iii) posttranslation proteolysis. On the basis of sequence analysis of the viral genome, translation starting at the first internal AUG codon within *gag* should yield the 34-kDa polypeptide (Fig. 3B).

The high-molecular-weight products, the putative Gag-Pro-Pol and Gag-Pro polyproteins, are produced in relatively small amounts compared with the Gag polypeptide (Fig. 2 and 3B). The reticulocyte cell-free system is relatively inefficient in translating high-molecular-weight proteins because of early quitting. The apparent ratio of these three translation products may be due to inefficiency of the reticulocyte translation system, or it may reflect an inherent property of the HTLV-2 mRNA which controls the ratio of these three products. To differentiate between these possibilities, a DNA fragment located between two unique *Sac*II endonuclease restriction sites was removed (Fig. 1), generating an internal deletion within the *gag* gene (pNM10A). Assuming that the three polypeptides initiate at the initiation codon of the *gag* ORF, the internal deletion within *gag* should similarly reduce the sizes of all three polypeptides. In fact, RNA transcribed from pNM10A linearized with *Hin*dIII in the polylinker (SAC-RNA) yielded the following three polyproteins with reduced molecular masses: Gag at 48 kDa, Gag-Pro at 63 kDa, and Gag-Pro-Pol at 96 kDa

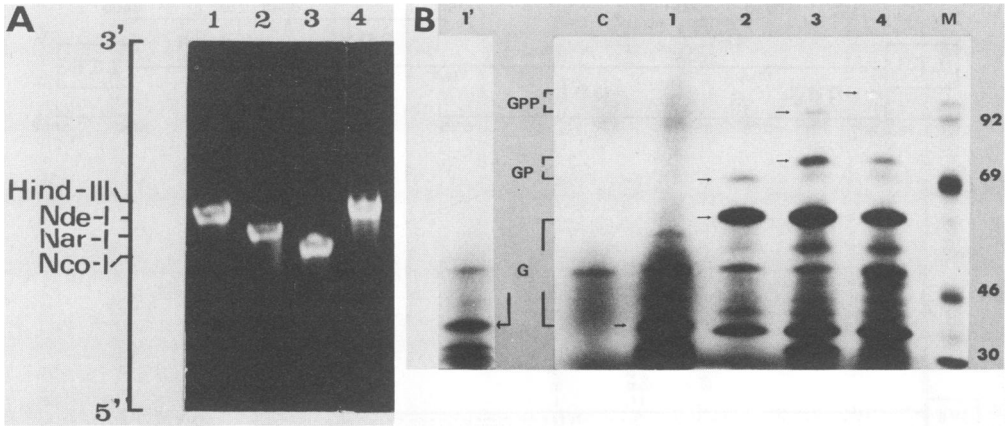

FIG. 3. Mapping of HTLV-2 gene products by translation of RNAs truncated at each of the three reading frames. (A) Electrophoresis on agarose gel (1%) of RNA species which were transcribed from pNM10 cleaved with the restriction endonucleases indicated on the left. Lanes: 1, *Nde*I; 2, *Nar*I; 3, *Nco*I; 4, *Hin*dIII. 28S and 18S rRNAs were used as molecular mass markers. (B) Translation products of the RNA species shown in panel A were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Lanes: 1 and 1', NCO-RNA (lane 1' illustrates the same autoradiogram but after a shorter exposure); 2, NAR-RNA; 3, NDE-RNA; 4, HIND-RNA; C, control translation without exogenous mRNA; M, molecular mass markers (sizes are indicated in kilodaltons on the right). The gene products and their truncated forms are indicated by arrows originating at G (Gag), GP (Gag-Pro), and GPP (Gag-Pro-Pol).

(Fig. 4, lane 2). The finding that the three polyproteins were similarly reduced in size because of the deletion in *gag* (compare lanes 1 and 2) indicates that these products start at the same initiation codon and confirms the identification of the HTLV-2 Gag, Gag-Pro, and Gag-Pro-Pol polyproteins. Furthermore, Fig. 4, lanes 1 and 2, and analysis of the autoradiograms by scanning (data not shown) showed that the internal deletion in *gag* did not increase the relative amounts of the Gag-Pro and Gag-Pro-Pol polyproteins. Thus, it may be concluded that differences in the relative amounts of the three polypeptides reflect an inherent property rather than the effect of distance from the initiation codon. Sequence analysis of HTLV-2 DNA indicated that synthesis of the Gag-Pro-Pol polyprotein requires two frameshift events, one between *gag* and *pro* and the other between *pro* and *pol*. We investigated whether readthrough at the *pro-pol* junction is dependent upon the first readthrough event at *gag-pro*. Cleavage of pNM10 with restriction endonuclease *Bsp*MII deleted a DNA fragment of 377 base pairs which comprises the *gag-pro* junction, including the *gag* translation termination codon (plasmid pNM10B; Fig. 5). This deletion joined the *gag* and *pro* ORFs to produce only two potential products; a 60-kDa Gag-Pro protein and a

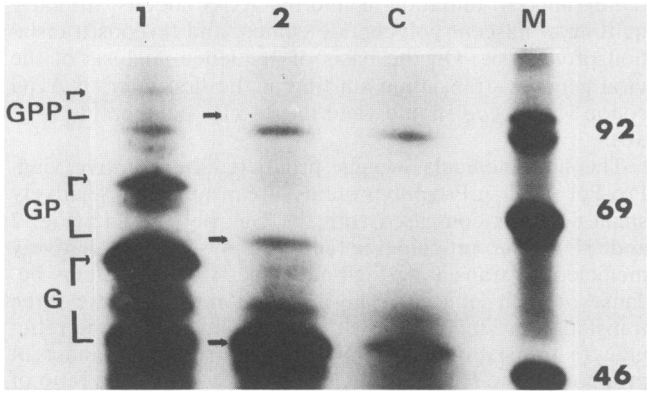

FIG. 4. Mapping of HTLV-2 gene products by translation of RNA deleted within the *gag* gene. Plasmid pNM10 was deleted of 309 base pairs within the *gag* gene by endonuclease *Sac*II (pNM10A) and linearized with endonuclease *Hin*dIII in the polylinker for transcription in the SP6 polymerase system (SAC-RNA). Lanes: 1, translation products of HIND-RNA (made with pNM10 [Fig. 2]); 2, translation products of SAC-RNA; C, control without exogenous mRNA; M, molecular mass markers (sizes are indicated in kilodaltons on the right). The decreases in the sizes of the polypeptides due to *Sac*II endonuclease deletion are indicated by arrows originating at Gag (G), Gag-Pro (GP), and Gag-Pro-Pol (GPP).

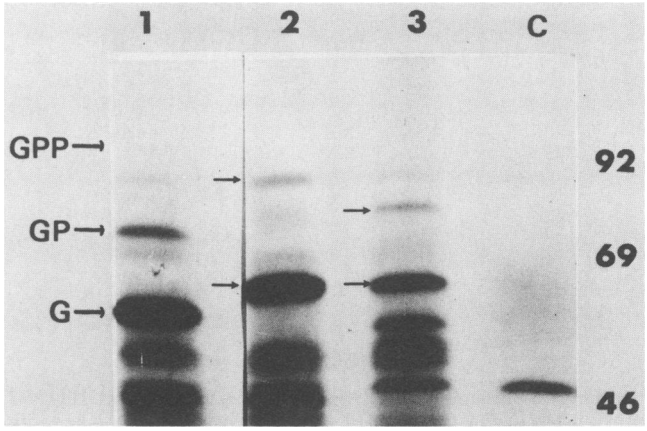

FIG. 5. Expression of *pol* gene in RNA deleted for the *gag-pro* junction. Plasmid pNM10 was cleaved twice by endonuclease *Bsp*MII to delete 377 base pairs at the *gag-pro* junction, including the *gag* translation termination codon (pNM10B [Fig. 1]). Plasmid pNM10B was linearized for transcription with endonuclease *Hin*dIII in the polylinker (BSP-RNA) or with endonuclease *Nde*I within the *pol* gene (BSP-RNA-NDE). The RNAs were translated, and the proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. The proteins Gag (G), Gag-Pro (GP), and Gag-Pro-Pol (GPP) made of HIND-RNA (plasmid pNM10) are indicated in lane 1. The corresponding internally deleted gene products Gag-Pro and Gag-Pro-Pol are made of BSP-RNA (lane 2) and BSP-RNA-NDE (lane 3) are indicated by arrows between the lanes. Lane C contained a control without exogenous mRNA.

104-kDa Gag-Pro-Pol readthrough product. In fact, translation of plasmid pNM10B linearized for transcription with *Hin*dIII in the polylinker (BSP-RNA) yielded two products of 61 and 91 kDa (Fig. 5, lane 2). This indicates that the second frameshift event is qualitatively independent of the first. Analysis of the autoradiogram presented in Fig. 5 was complicated, since the 91-kDa protein comigrated with an endogenous reticulocyte translation product (Fig. 5, faint band in lane C). To overcome this problem, pNM10B DNA was cleaved within the *pol* gene with restriction endonuclease *Nde*I (Fig. 1). As expected, translation of plasmid pNM10B linearized with *Nde*I within the *pol* gene (BSP-RNA-NDE) yielded a 61-kDa Gag-Pro product and a polyprotein of 81 kDa (lane 3) which replaced the 91-kDa Gag-Pro-Pol product (lane 2).

The frequencies of readthrough may potentially be dictated by rates of translation at the region proximal to the frameshift site. On the basis of this assumption, it is expected that translation of BSP-RNA-NDE deleted of the *gag pro* junction would produce elevated levels of the *pol* product compared with those produced by HIND-RNA, in which only 10 to 12% of the ribosomes migrated from *pro* to *pol* (Fig. 2). Indeed, scanning of autoradiograms at various exposure times in repeated experiments indicated about 27% readthrough at the *pro-pol* junction of BSP-RNA-NDE (Fig. 5). These results suggest that readthrough efficiency is affected to some degree by translation rates at the region upstream of the ORF junction.

Is there a consensus signal for the two translation frameshifts in HTLV-2? The sequence at the junction of *gag-pro* in MMTV, HTLV-1, HTLV-2, and bovine leukemia virus is characterized by the presence of a consensus sequence, AAAAAAC, adjacent to the UAA termination codon (2, 6, 14, 18). On the other hand, no common features are apparent at the overlap of the *pro* and *pol* genes of these four viruses; any of the three termination codons separate these two ORFs, and a consensus sequence, UUUA, is found adjacent to the termination codon in three of these viruses. In HTLV-1 a similar sequence is located 156 nucleotides upstream of the termination codon (4). Although no obvious common signals for the two frameshift events in HTLV-2 were observed, our work indicates that similar readthrough frequencies are dictated at the two junctions when they are located in tandem.

In some viruses, such as Rous sarcoma virus, a sequence with a potential of forming an RNA stem-and-loop structure was shown to be involved in the frameshift event (4). However, in other viruses, such as human immunodeficiency virus, it was demonstrated that frame shifting is not dependent on stem-and-loop structures (19). In HTLV-2, the two different homopolymeric sequences described above are followed by a sequence with a potential of forming a stem-and-loop structure. Thus, HTLV-2 presents an experimental system for analysis of the role of such structures in the two frameshift events.

The following three strategies are used by retroviruses for expression of the *pol* gene: (i) occasional insertion of a glutamine at the nonsense codon, as in murine leukemia virus (20); (ii) changing a reading frame in Rous sarcoma virus and human immunodeficiency virus (4, 5, 7, 15); (iii) shifting twice the reading of translation frames in MMTV (6) and HTLV-2. Although different readthrough mechanisms are used by Rous sarcoma virus and murine leukemia virus, the relative amounts of the Pol and Gag proteins produced in vitro and in chronically infected cells are similar at 5 to 10% (4, 7, 12). Our in vitro experiments with HTLV-2 RNA indicate that two translation frameshift events are required for the synthesis of Pol, and therefore the relative amounts of Pol and Gag are only about 1:100. It would be of interest to determine whether this ratio also holds in virus-infected cells.

We thank H. Falk for excellent technical assistance, A. Maschler for preparation of the manuscript, and M. Aboud for the HTLV-2 DNA clone.

This work was supported by a fellowship from the Israel Cancer Research Fund and by the Dr. Karl and Dr. Lea Terner Foundation for Cancer Research.

LITERATURE CITED

1. **Hattori, S., T. Kiyokawa, K. Imagawa, F. Shimizu, E. Hashimura, M. Seiki, and M. Yoshida.** 1984. Identification of gag and env gene products of human T-cell leukemia virus (HTLV). Virology **136**:338–347.
2. **Hiramatsu, K., J. Nishida, A. Naito, and H. Yoshikura.** 1987. Molecular cloning of the closed circular provirus of human T-cell leukemia virus type I: a new open reading frame in the gag-pol region. J. Gen. Virol. **68**:213–218.
3. **Hizi, A., L. E. Henderson, T. D. Copeland, R. C. Sowder, C. V. Hixson, and S. Oroszlan.** 1987. Characterization of mouse mammary tumor virus gag-pro gene products and the ribosomal frameshift site by protein sequencing. Proc. Natl. Acad. Sci. USA **84**:7041–7045.
4. **Jacks, T., H. D. Madhani, F. R. Masiarz, and H. E. Varmus.** 1988. Signals for ribosomal frameshifting in the Rous sarcoma virus gag-pol region. Cell **55**:447–458.
5. **Jacks, T., M. D. Power, F. R. Masiarz, P. A. Luciw, P. J. Barr, and H. E. Varmus.** 1988. Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. Nature (London) **331**: 280–283.
6. **Jacks, T., K. Townsley, H. E. Varmus, and J. Majors.** 1987. Two efficient ribosomal frameshifting events are required for synthesis of mouse mammary tumor virus gag-related polyproteins. Proc. Natl. Acad. Sci. USA **84**:4298–4302.
7. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science **230**:1237–1242.
8. **Kalyanaraman, V. S., R. Narayanan, P. Feorino, R. B. Ramsey, E. L. Palmer, T. Chorba, S. McDougal, J. P. Getchell, B. Holloway, A. K. Harrison, C. D. Cabradilla, M. Telfer, and B. Evatt.** 1985. Isolation and characterization of a human T-cell leukemia virus type II from a hemophilia-A patient with pancytopenia. EMBO J. **4**:1455–1460.
9. **Levin, J. G., S. C. Hu, A. Rein, L. I. Messer, and B. I. Gerwin.** 1984. Murine leukemia virus mutant with frameshift in the reverse transcriptase coding region: implications for *pol* gene structure. J. Virol. **51**:470–478.
10. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green.** 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7056.
11. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61**:480–490.
12. **Murphy, E. C., J. J. Kopchick, K. F. Watson, and R. B. Arlinghaus.** 1978. Cell-free synthesis of a precursor polyprotein containing both gag and pol gene products by Rauscher murine leukemia virus 35S RNA. Cell **13**:359–369.
13. **Ratner, L., S. F. Josephs, B. Starcich, B. Hahn, G. M. Shaw, R. C. Gallo, and F. Wong-Staal.** 1985. Nucleotide sequence analysis of a variant human T-cell leukemia virus (HTLV-Ib) provirus with a deletion in pX-I. J. Virol. **54**:781–790.
14. **Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden.** 1985. The gag and pol genes of bovine leukemia virus: nucleotide sequence and analysis. Virology **142**:357–377.
15. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.

16. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
17. **Shaw, G. M., M. A. Gonda, G. H. Flickinger, B. H. Hahn, R. C. Gallo, and F. Wong-Staal.** 1984. Genome of evolutionarily divergent members of the human T-cell leukemia virus family (HTLV-I and HTLV-II) are highly conserved, especially in pX. Proc. Natl. Acad. Sci. USA **81**:4544–4548.
18. **Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura.** 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA **82**:3101–3105.
19. **Wilson, W., M. Braddock, S. E. Adams, P. D. Rathjen, S. H. Kingsman, and A. J. Kingsman.** 1988. HIV expression strategies: ribosomal frameshifting is directed by a short sequence in both mammalian and yeast systems. Cell **55**:1159–1169.
20. **Yoshinaka, Y., I. Katoh, T. D. Copeland, and S. Oroszlan.** 1985. Murine leukemia virus protease is encoded by the gag-pol gene and is synthesized through suppression of an amber termination codon. Proc. Natl. Acad. Sci. USA **82**:1618–1622.