JOURNAL OF VIROLOGY, Jan. 1993, p. 196–203
0022-538X/93/010196-08$02.00/0

Vol. 67, No. 1

Characterization of Ribosomal Frameshifting for Expression of *pol* Gene Products of Human T-Cell Leukemia Virus Type I

SEOK HYUN NAM,[1]† TERRY D. COPELAND,[1] MASAKAZU HATANAKA,[2] AND STEPHEN OROSZLAN[1]*

Laboratory of Molecular Virology and Carcinogenesis, ABL-Basic Research Program, National Cancer Institute-Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201,[1] and Institute for Virus Research, Kyoto University, Sakyo-ku, Kyoto 606, Japan[2]

Received 28 February 1992/Accepted 5 October 1992

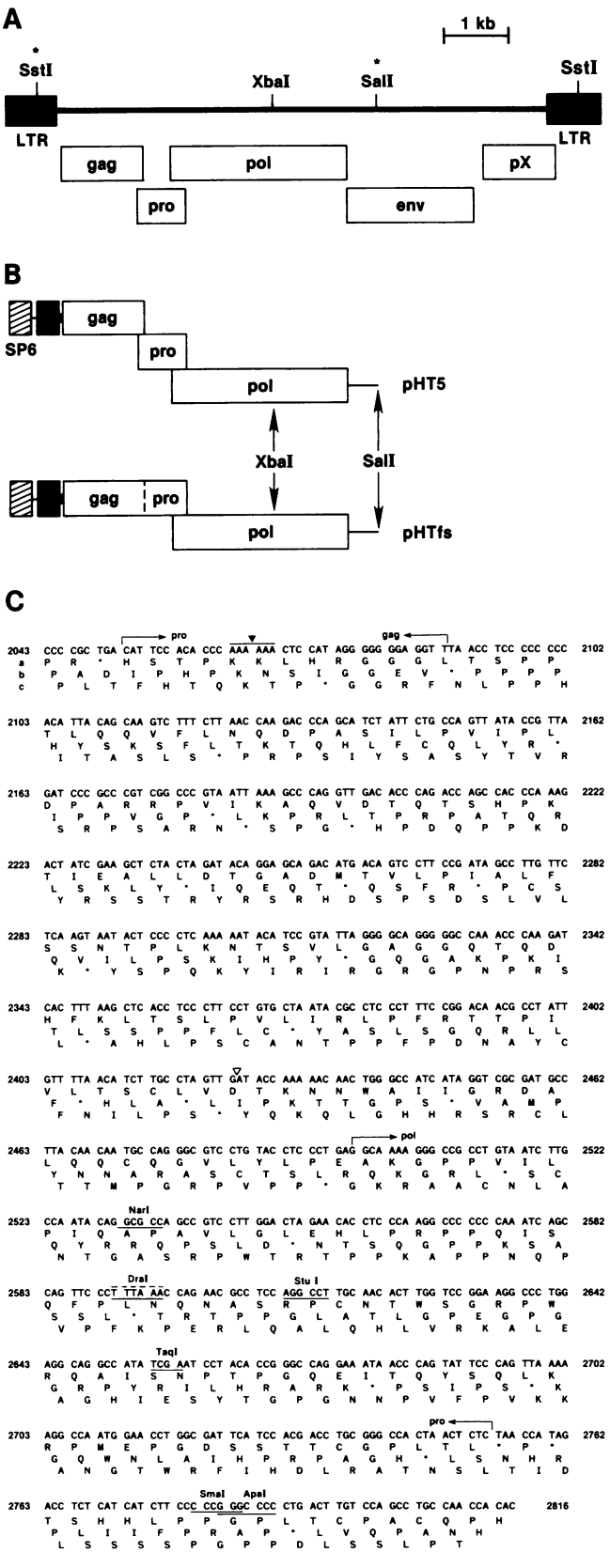
For study of the *pol* gene expression of human T-cell leukemia virus type I (HTLV-I), RNA was transcribed in vitro from proviral DNA and translated in rabbit reticulocyte lysates. This cell-free translation resulted in two major translation products representing the Gag and Gag-Pro polyproteins. By contrast, the Gag-Pro-Pol polyprotein could be readily observed only when translation was performed with mutant mRNA in which the protease (*pro*) reading frame was aligned to *gag* to eliminate the frameshifting event in the *gag-pro* overlap. The results indicated that two independent ribosomal frameshifting events are required for expression of the HTLV-I *pol* gene product. Studies with mutant DNAs facilitated the characterization of the primary structure of the HTLV-I mRNA responsible for the ribosomal frameshift in the *pro-pol* overlap and demonstrated that the frameshift occurs at the signal sequence UUUAAAC. Direct amino acid sequencing of the transframe protein localized the site of the frameshift to the asparagine codon AAC.

Human T-cell leukemia virus type I (HTLV-I), the first human exogenous retrovirus, was found to be etiologically associated with human adult T-cell leukemia (8, 24, 25, 38). As is true for all replication-competent retroviruses, the HTLV-I genome has the three genes, *gag*, *pol*, and *env*, that are known to be essential for the production of infectious progeny virus. In mammalian type C viruses, the *gag* and *pol* open reading frames (ORFs) are in phase but are separated by an amber termination codon (33). In other retroviruses, the *gag* and *pol* ORFs are out of phase and the 5′ end of the *pol* ORF overlaps the 3′ end of the *gag* ORF. The protease (PR) gene (*pro*) may be located at either the 3′ end of *gag* or the 5′ end of *pol*. Respective examples are human immunodeficiency virus type 1 (HIV-1) (27) and avian Rous sarcoma virus (RSV) (31). However, in the HTLV family of retroviruses, including bovine leukemia virus, as well as in mouse mammary tumor virus (MMTV) and type D retroviruses, the *gag* and *pol* genes are separated by a *pro* ORF that overlaps both *gag* and *pol* (13, 20, 22, 26, 28, 29, 32, 34).

The internal structural proteins are encoded by the *gag* gene and are designated the matrix (MA), capsid (CA), and nucleocapsid (NC) proteins (16). Viral PR is encoded by *pro*, and reverse transcriptase (RT) and integrase (IN) are derived from the *pol* gene. All of the above-described mature virion components are generated by proteolytic cleavage of the primary translational products of the genome-sized mRNA, designated the Gag, Gag-Pro, Gag-Pol, and Gag-Pro-Pol polyproteins. The synthesis of Gag-Pro, Gag-Pol, and Gag-Pro-Pol fusion proteins, which all share the common Gag initiator codon located at the 5′ end of the *gag* gene, is translationally regulated. Two general mechanisms by which the synthesis of the fusion proteins can be achieved are known: in-frame readthrough and ribosomal frameshifting. In mammalian type C viruses, the *gag* amber terminator (UAG) is suppressed by a glutamine-tRNA to translate the *pol* gene (39). In all other retroviruses, ribosomal frameshifting in the −1 direction is required to align the different ORFs (for a review see references 6, 7, and 10). Depending on the genomic organization of the particular virus, either a single frameshift or two frameshifts, one in the *gag-pro* overlap and the other in the *pro-pol* overlap, are required to synthesize the *pol*-derived replication enzymes. The site of the frameshifting between *gag* and *pol* has been determined by site-directed mutagenesis and amino acid sequencing of the in vitro-synthesized transframe peptides of RSV (11) and HIV (12), which are examples for the single frameshift. MMTV is an example for the double frameshift (13, 20). In addition to the mutational studies of Jacks et al. (13), sequencing of the natural viral transframe protein has determined the first frameshift site (*gag-pro*) of MMTV (9). Nam et al. (23) reported previously that the Gag-Pro fusion protein of HTLV-I is also translated via ribosomal frameshifting in a vaccinia virus expression vector system. Site-specific mutagenesis experiments showed that the six consecutive adenine residues found in the *gag-pro* overlap are involved in the slippage of tRNA to the −1 *pro* frame to generate the Gag-Pro fusion protein that is the precursor of the PR (23). The HTLV-I *pol* ORF is also −1 relative to the upstream *pro* ORF (22). As expected on the basis of the genomic organization and studies of HTLV-II (17), two frameshifting events are required for HTLV-I *pol* expression. To date, the second frameshift site has not been determined in HTLV-I or in any other retrovirus by use of both site-directed mutagenesis and sequencing of the transframe protein (9).

In this study, to test how and where the second frameshifting event occurs in the *pro-pol* overlap of HTLV-I, we have used an approach similar to that used to demonstrate frameshifting in RSV (11). DNAs corresponding to the entire *gag*, *pro*, and *pol* coding regions and appropriate mutant DNAs were transcribed in vitro with bacteriophage SP6

* Corresponding author.
† Present address: Department of Biology, Ajou University, Suwon 441-749, Republic of Korea.



RNA polymerase into the respective RNAs. Each RNA was translated in a cell-free translation system, and the resulting translational products were analyzed by immunoprecipitation and then by sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis. The results obtained support the idea that the *pol* gene product of HTLV-I is synthesized as a fusion protein with Gag-Pro via a second frameshifting event occurring in the *pro-pol* overlap. Radiolabeled amino acid sequencing of the transframe peptide and in vitro mutagenesis were used to localize the precise site of frameshifting to the 3′ end of an AAC asparagine codon. The asparagine codon is located in the sequence U UUA AAC, which occurs in the 5′ region of the *pro-pol* overlap of HLTV-I.

MATERIALS AND METHODS

General DNA methods. Plasmid DNA was purified by a slight modification of the alkali-SDS method of Birnboim and Doly (1). Restriction endonucleases and DNA-modifying enzymes were used as specified by the manufacturers. The enzymes used and their sources were as follows: *Xba*I, *Kpn*I, *Xma*III, *Bsu*36I, T4 DNA polymerase, *Sal*I, Klenow fragment of *Escherichia coli* DNA polymerase, *Nar*I, *Stu*I, *Hin*dII, *Hin*dIII, *Dra*I, *Apa*I, and *Xho*I were obtained from New England BioLabs, and *Sst*I, *Ssp*I, and *Sma*I were obtained from Bethesda Research Laboratories. All molecular biological manipulations were carried out by standard methods (18). All DNA transformations were performed with competent *E. coli* JM109 cells. However, single-stranded DNAs for site-specific mutagenesis were produced by use of transformants of *E. coli* TG1 harboring subcloned M13 replicative-form (RF) DNAs. The sequencing of single-stranded M13 virion DNA or double-stranded RF DNAs was done by the chain termination method described by Sanger et al. (30) with a ^{32}P-DNA sequence kit purchased from New England BioLabs.

Plasmid construction and site-specific mutagenesis. To construct plasmid pHT5 (Fig. 1), used for runoff transcription, we isolated the 5.1-kb *Sst*I-*Kpn*I fragment of clone HTLV1C, which contains the 3′ half of the long terminal repeat, *gag*, *pro*, *pol*, and the 5′ portion of *env* (21). The

FIG. 1. Construction of the recombinant clones used for in vitro frameshifting. (A) Organization of the HTLV-I genome (see references 22 and 32). ORFs are drawn below the diagram of the proviral DNA. The closed boxes represent the long terminal repeats (LTR). The *Sst*I and *Sal*I sites marked with asterisks were used to excise the region harboring the entire *gag*, *pro*, and *pol* ORFs. (B) Diagram of genomic arrangements of recombinant DNA inserted into pSP65 for in vitro transcription (19). Insertion of the wild-type DNA fragment downstream of the SP6 promoter produced pHT5. Mutant pHTfs was created by a 1-base insertion into the *gag-pro* overlap region, as shown by the vertical broken line, to align these ORFs. Either the *Xba*I or the *Sal*I site (as shown) was used to linearize these plasmids for in vitro transcription. (C) Nucleotide sequence and translated amino acid sequence of the HTLV-I *gag-pro* and *pro-pol* overlapping regions (22, 32). (a) *pro*. (b) *gag*. (c) *pol*. Arrows show the beginning and end of the ORFs in pHTfs. The site of the single point mutation A → G, used to create a new *Kpn*I restriction site is indicated by an open triangle. Solid lines under the nucleotide sequence show the restriction sites, as indicated. The solid line above the nucleotide sequence shows the signal sequence for frameshifting in the *gag-pro* overlap. Above this line, the closed triangle shows the site of the 1-base insertion that aligns the *gag* and *pro* ORFs in pHTfs. The broken line above the nucleotide sequence shows the signal sequence in the *pro-pol* overlap, as determined in this report.

purified DNA fragment was ligated with *Sst*I-*Sal*I-digested pSP65 (19), and then plasmid DNA from an ampicillin-resistant transformant was tested to select the desired plasmid by appropriate enzyme digestion.

To construct pHTfs (Fig. 1), we digested pHT5 to completion with *Xma*III and *Bsu*36I. The 6.8-kb *Xma*III-*Bsu*36I fragment harboring the vector sequence was isolated from plasmid p7.5gagfs19AspII, described in a previous report (23).

Oligonucleotide-directed in vitro mutagenesis was performed as described by Taylor et al. (35) with a kit purchased from Amersham. We isolated the 0.59-kb *Kpn*I-*Pst*I fragment including a portion of the *pro-pol* overlap from pHTfs and then inserted it into the polylinker site of M13 RF DNA. The recombinant RF DNA was used to prepare the single-stranded DNA to be used as a template for creation of the following mutations. One is a base substitution from A to G to convert the AAC asparagine codon (nucleotides [nt] 2595 to 2597) to an AGC serine codon. M13 RF DNA harboring the point mutation was digested with *Bsu*36I and *Apa*I to isolate the 0.29-kb fragment. Then, this fragment was ligated with the 7.9-kb *Bsu*36I-*Apa*I fragment from pHTfs to obtain the plasmid designated pHTfsΔFG. The same experimental procedure was used to construct another plasmid, designated pHTFsCG, in which, instead of a single base change, the hexanucleotide CCAGAA (nt 2597 to 2602) was substituted with the sequence of another hexanucleotide, CTGCAG. The construction of other plasmids derived from pHT5 and pHTfs is described in the figure legends.

Preparation of the DNA template and transcription into mRNA. Plasmids for transcription were prepared by lysis of plasmid-containing bacteria with Triton X-100 and purified by double banding on a cesium chloride-ethidium bromide density gradient (18). In vitro transcriptions were carried out essentially as described by Melton et al. (19), with a modification to generate capped mRNA (15). mRNA was synthesized in vitro with a capping kit from Stratagene, except that SP6 RNA polymerase and polymerase buffer were purchased from Promega Biotec. Following the transcription reaction at 40°C for 60 min, the mixture was extracted once with phenol-chloroform (1:1 [vol/vol]) and once with chloroform, and the mRNA was ethanol precipitated twice in the presence of 0.8 M ammonium acetate. RNA was dissolved in water and checked for integrity by electrophoresis on 1% agarose–formaldehyde gels (18).

Cell-free translation of mRNA and immunoprecipitation. Rabbit reticulocyte lysates were purchased from Promega, and reactions were performed under the conditions specified by the manufacturer. In brief, protein synthesis was carried out in a 50-μl reaction volume containing 1 μg of mRNA and 1 μCi of [^{35}S]methionine (Amersham) per μl. After incubation for 60 min at 30°C, [^{35}S]methionine-labeled viral proteins were immunoprecipitated with monoclonal antibody to p19^{gag} or with the appropriate rabbit antisera. The translation mixture and antiserum (or monoclonal antibody), together with protein A-Sepharose (Pharmacia LKB Biotechnology), were incubated for 16 h at 4°C in 20 mM Tris-HCl (pH 7.2)–50 mM NaCl–0.5% sodium deoxycholate–0.5% Nonidet P-40. The beads were washed three times with the same buffer and once with ice-cold distilled water, and the immunoprecipitates were subjected to electrophoresis in SDS–10% polyacrylamide gels.

Synthesis, purification, and sequencing of the transframe protein. To construct plasmid pHA, we excised the 1.4-kb *Pst*I-*Sma*I fragment containing the viral p24^{gag}, p15^{gag}, and PR coding sequences from pHT5. It was then ligated with pUC18 that had been linearized with *Pst*I and *Sma*I. The recombinant plasmid was partially digested with *Dra*I and then completely digested with *Hin*dIII to obtain the 2.9-kb fragment. This fragment was ligated with the annealed double-stranded synthetic linker

```
3'-ATACTCCAGGGATCCCAAGGGAAA-5'
5'-AGCTTATGAGGTCCCTAGGGTTCCCTTT-3'
```

to introduce the specific leader sequence encoding the initiator methionine plus 4 amino acids derived from chick prelysozyme protein upstream of the putative *pro-pol* frameshift site. This amino acid sequence has been known to prevent the acetylation of the amino termini of proteins during in vitro translation (12 and references therein). The resultant plasmid was completely digested with *Sma*I and *Hin*dIII to obtain the 0.22-kb *Hin*dIII-*Sma*I fragment containing the leader sequence. This fragment was ligated with pSP64 that had been linearized with the corresponding enzymes. Then, we isolated the 1.2-kb *Staphylococcus aureus* protein A gene fragment from pRIT2T (Pharmacia) by complete digestion with *Hin*dIII and *Pvu*II (37). Following treatment with the Klenow fragment to fill in the ends, the protein A gene fragment was inserted into the *Sma*I site of pSP64 harboring the *pro-pol* overlap plus the linker sequence between the *Hin*dIII and *Sma*I sites. In plasmid pHA, the protein A gene segment is continuous with the *pol* frame of the HTLV-I insert.

Cell-free translation with a rabbit reticulocyte lysate was performed with mRNA generated by runoff transcription from pHA as described above, except that the translation product was radiolabeled with [^{35}S]methionine in combination with [^{3}H]leucine, [^{3}H]proline, [^{3}H]phenylalanine, or [^{3}H]lysine (Amersham) in the presence of 25 μg of α_2-macroglobulin (Boehringer Mannheim Biochemicals) per ml. Subsequently, the chimeric proteins synthesized in vitro were purified with rabbit immunoglobulin G-Sepharose (Pharmacia) by the method specified by the manufacturer. Radioactive fractions eluted with 0.5 M acetic acid (pH 3.4) were collected and then lyophilized. The purified proteins were subjected to 25 cycles of Edman degradation in a Beckman model C liquid-phase sequencer as described previously (40), with myoglobin as a carrier.

Antibodies. To detect the Gag-related polyproteins synthesized in the cell-free translation system, we used a monoclonal antibody against the p19^{gag} protein, termed GIN7, as previously described (23). Antisera against synthetic peptides SP-102 (*pro* ORF) and SP-297 (*pol* ORF, RNase H region) were also used. SP-102 is a decapeptide (H-GlyLeuGluHisLeuProArgProProGlu-NH_2), and SP-297 is a nonadecapeptide (H-CysSerGlnArgSerPheProLeuProPro ProHisLysSerAlaGlnArgAlaGlu-NH_2). Both peptides were conjugated to keyhole limpet hemocyanin for immunization of rabbits.

RESULTS

Characterization of Gag-related polyproteins expressed in a cell-free translation system. In a previous study (23), it was reported that HTLV-I PR is synthesized as part of a Gag-Pro precursor polyprotein by ribosomal frameshifting at six consecutive adenine residues within the *gag-pro* overlap window and subsequently processed into mature PR. We anticipated that a similar translational mechanism is necessary to synthesize the *pol* gene products RT and IN, because the *pol* gene overlaps the *pro* gene in the −1 direction and there is no initiation methionine codon within the *pro-pol*

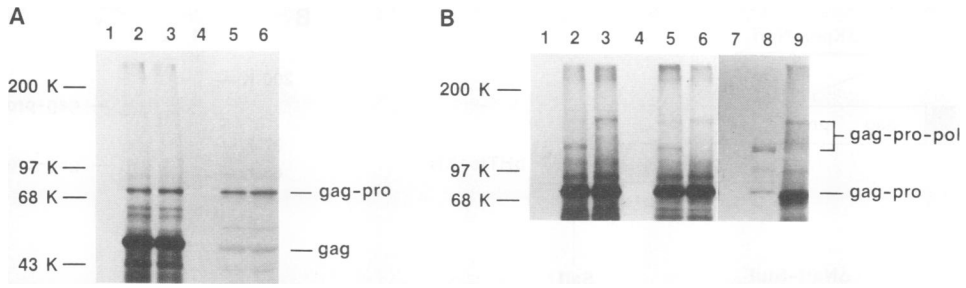


FIG. 2. SDS-polyacrylamide gel electrophoresis of cell-free translation products directed by the in vitro-synthesized RNAs. (A) Reticulocyte lysate translation products synthesized in response to RNA transcribed from *Xba*I-digested pHT5 (lanes 2 and 5) RNA transcribed from *Sal*I-digested pHT5 (lanes 3 and 6), or no RNA (lanes 1 and 4). RNA was translated and products were labeled with [^{35}S]methionine as described in Materials and Methods. Polypeptides were immunoprecipitated with an anti-p19^{gag} monoclonal antibody (lanes 1, 2, and 3) or a rabbit antiserum raised against a synthetic peptide corresponding to a PR region (lanes 4, 5, and 6). The positions of the Gag (53-kDa) and Gag-Pro (76-kDa) proteins and molecular mass markers (in thousands, on the left) are indicated. (B) Fluorogram of [^{35}S]methionine-labeled proteins synthesized in response to RNA transcribed from *Xba*I-digested pHTfs (lanes 2, 5, and 8), RNA transcribed from *Sal*I-digested pHTfs (lanes 3, 6, and 9), or no RNA (lanes 1, 4, and 7). Immunoprecipitation was performed with an anti-p19^{gag} monoclonal antibody (lanes 1, 2, and 3), with a rabbit antiserum raised against a synthetic peptide corresponding to a *pro* ORF region (lanes 4, 5, and 6), or a rabbit antiserum raised against a synthetic peptide corresponding to the middle region of the RNase H domain (lanes 7, 8, and 9). The positions of the Gag-Pro (76-kDa) and Gag-Pro-Pol (115- and 160-kDa) proteins and molecular mass markers are indicated. The nature of the unexpected strong band that was coprecipitated with the largest translational product (160 kDa) and migrated as a 70-kDa protein (lane 9) is not known. See the text for a possible explanation.

overlap (22). Moreover, we observed the presence within this overlap of the nucleotide sequence CCCUUUAAAC, containing both UUUA and AAAC, which have been predicted and/or found to be the frameshifting sites in several retroviruses (9–14, 20, 23, 28). To test whether the HTLV-I *pol* gene is expressed by ribosomal frameshifting, we introduced into the downstream region of the SP6 promoter the 5.1-kb *Sst*I-*Sal*I proviral DNA fragment derived from infectious clone HTLVIC (21). This fragment harbors the entire *gag*, *pro*, and *pol* genes and the 5′ portion of the *env* gene. The relevant regions of this plasmid, pHT5, are outlined in Fig. 1 (see also Materials and Methods). pHT5 was completely digested with *Xba*I or *Sal*I, and the linearized plasmid was used as a template to synthesize RNA. In vitro transcription by SP6 RNA polymerase made an abundant amount of a single species of RNA, either 3.6 or 5.1 kb in length, in which the HTLV-I *gag*, *pro*, and *pol* coding regions were in their genomic configuration.

The in vitro-synthesized 5.1- or 3.6-kb RNA was added to a rabbit reticulocyte cell-free translation system. Typical [^{35}S]methionine-labeled translation products of each RNA are shown in Fig. 2A. Fractionation by SDS-polyacrylamide gel electrophoresis revealed two major translation products, 53- and 76-kDa polyproteins. Both were immunoprecipitated with an anti-p19^{gag} monoclonal antibody (Fig. 2A, lanes 1 to 3). However, the synthetic peptide (SP-102) antiserum specific for the *pro* ORF product of HTLV-I recognized only the 76-kDa protein (Fig. 2A, lanes 4 to 6). As previously found for the vaccinia virus expression vector system (23), the 53-kDa product is the Gag precursor polyprotein, and the 76-kDa product is the Gag-Pro fusion protein synthesized by ribosomal frameshifting. The nature of the minor reaction products observed when the anti-p19^{gag} monoclonal antibody was used is not known. They may represent products of internal initiation and/or premature termination.

The Gag-Pro-Pol fusion protein, which is predicted to have a molecular mass of 160 kDa, could not be observed in our cell-free translation system charged with the RNAs derived from pHT5, in which the *gag*, *pro*, and *pol* genes are in their genomic configuration. The failure to detect this polyprotein was probably due to the low efficiency of the second consecutive frameshifting required to translate *pol*. We thought that if the expression of the HTLV-I *pol* gene product also depended on ribosomal frameshifting within the *pro-pol* overlap, a larger amount of the Gag-Pro-Pol precursor polyprotein should be made when a mutagenized RNA in which the *gag* and *pro* genes are aligned to allow synthesis of the largest fusion protein (160 kDa) by a single instead of a double frameshift is used. Therefore, we constructed plasmid pHTfs for RNA runoff transcription. This plasmid has the same molecular structure as pHT5, except that a single base insertion was created at the consecutive adenine residues within the *gag-pro* overlap to align these ORFs (Fig. 1B and C). As with pHT5, restriction of pHTfs with *Sal*I or *Xba*I and subsequent transcription generated a 5.1- or 3.6-kb RNA, respectively. As expected, translation of the RNAs in the rabbit reticulocyte lysate system produced the 76-kDa protein as the major product and two minor translational products of 160 and 115 kDa, depending on which endonuclease was used for restriction. All of these proteins were immunoprecipitable with the anti-p19^{gag} monoclonal antibody (Fig. 2B, lanes 2 and 3) and a rabbit antiserum to a synthetic decapeptide, SP-102 (Fig. 2B, lanes 5 and 6), which is derived from the *pro* ORF. The minor translational products, the proteins of 160 and 115 kDa, were also immunoprecipitable when we used an antiserum raised against a synthetic peptide (SP-207) corresponding to an RNase H region (Fig. 2B, lanes 8 and 9). The nature of the unexpected strong band of ~70 kDa in lane 9 is not known. It is likely the product of internal initiation in the *pol* ORF. This idea is supported by the appearance of a ~25-kDa protein in Fig. 2B, lane 8 (data not shown). These results show that as for HTLV-I *pro* gene expression (23), ribosomal frameshifting in the −1 direction is also operative within the *pro-pol* overlap and allows the synthesis of the 160-kDa Gag-Pro-Pol precursor polyprotein. The 160-kDa protein should result from translation extending into the HTLV-I *pol* region and should possess the RT domain as well as the IN domain. This idea is supported by the fact that an antiserum against a synthetic peptide corresponding to the carboxyl-terminal portion of

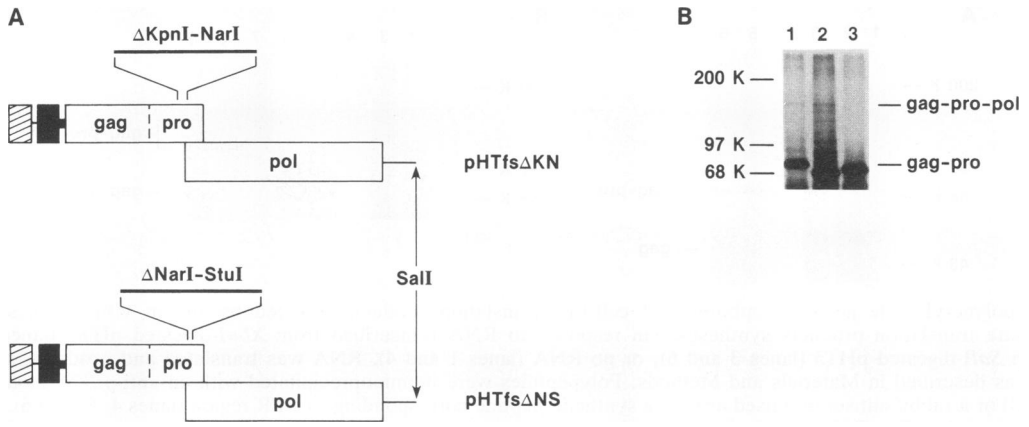


FIG. 3. Analysis of deletion mutants of pHTfs for mapping the RNA region responsible for the ribosomal frameshifting within the *pro-pol* overlap. (A) Schematic representation of deletion plasmids pHTfsΔKN and pHTfsΔNS. These were constructed from pHTfs by restriction endonuclease digestion as indicated. The recognition sites of the respective endonucleases are shown in Fig. 1C. Plasmid pHTfsΔKN has a deletion of the nucleotide sequences between the *Kpn*I (Fig. 1C, open triangle) and *Nar*I sites. In pHTfsΔNS, the nucleotide sequences between the *Nar*I and *Stu*I sites, which involve the putative signal sequence, are deleted. (B) Fluorogram of ^{35}S-labeled proteins produced from rabbit reticulocyte lysate translation of RNAs transcribed from *Sal*I-digested pHTfsΔKN (lane 2) and pHTfsΔNS (lane 3). The RNA transcribed from *Sal*I-digested pHTfs was also used to synthesize protein (lane 1) in a control experiment. All polypeptides were immunoprecipitated with an anti-p19^{gag} monoclonal antibody. The positions of the Gag-Pro and Gag-Pro-Pol proteins and molecular mass markers are indicated.

HTLV-I IN also immunoprecipitated the same protein (data not shown).

Analysis of mutants for defining the frameshifting signal within the *pro-pol* overlap. To define the region of RNA necessary for ribosomal frameshifting, we constructed two deletion mutants derived from pHTfs (Fig. 3A). In plasmid pHTfsΔKN, the deletion of 108 nt between the *Kpn*I and *Nar*I sites resulted in the removal of 12 codons from the 5′ portion of *pol* as well as 24 codons from the *pro* ORF. Plasmid pHTfsΔNS was constructed by deleting 78 nt between the *Nar*I site and the *Stu*I site in the *pro-pol* overlapping region. A deletion spanning the region from nt 2424 to 2532 in pHTfsΔKN had no influence on proper ribosomal frameshifting. Cell-free translation of this mutant RNA produced about the same amount of *pol* gene product as that of pHTfs and, as expected, the mutant protein was somewhat smaller (Fig. 3B, lanes 1 and 2). In contrast, ribosomal frameshifting was completely blocked when the deletion occurred between nt 2532 and 2611 (plasmid pHTfsΔNS). We could no longer observe the larger protein molecule (Fig.

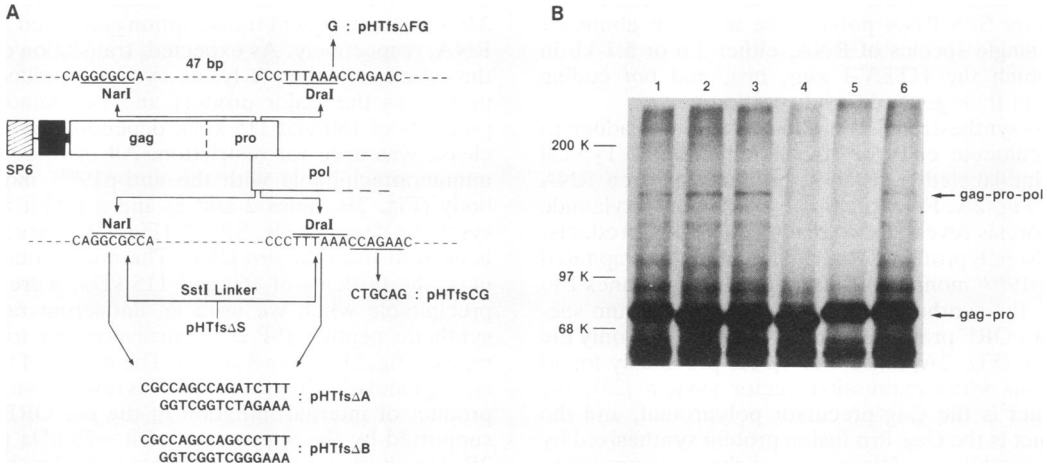


FIG. 4. DNA constructs carrying deletions or point mutations in the *pro-pol* overlap of pHTfs and their products. (A) Mutant clones were as follows. In pHTfsΔFG, the third A of the *Dra*I site is substituted with a G. In pHTfsCG, the wild-type sequence CCAGAA (underlined) is changed to CTGCAG. In pHTfsΔS, the wild-type sequence between the *Nar*I and *Dra*I sites is deleted, and an *Sst*I linker, the synthetic octanucleotide 5′-CGAGCTCG-3′, is introduced. In pHTfsΔA and pHTfΔB, double-stranded synthetic oligonucleotides with the sequences shown were introduced in place of the deleted sequences between the *Nar*I and *Dra*I sites. (B) Translation products synthesized in response to mRNAs derived from *Sal*I-digested pHTfs (lane 1), pHTfsΔB (lane 2), pHTfsΔA (lane 3), pHTfsΔS (lane 4), pHTfsΔFG (lane 5), and pHTfsΔCG (lane 6). The proteins labeled with [^{35}S]methionine were immunoprecipitated with a anti-p19^{gag} monoclonal antibody and separated on a 10% SDS–polyacrylamide gel. The positions of the Gag-Pro and Gag-Pro-Pol proteins and molecular mass markers are indicated.

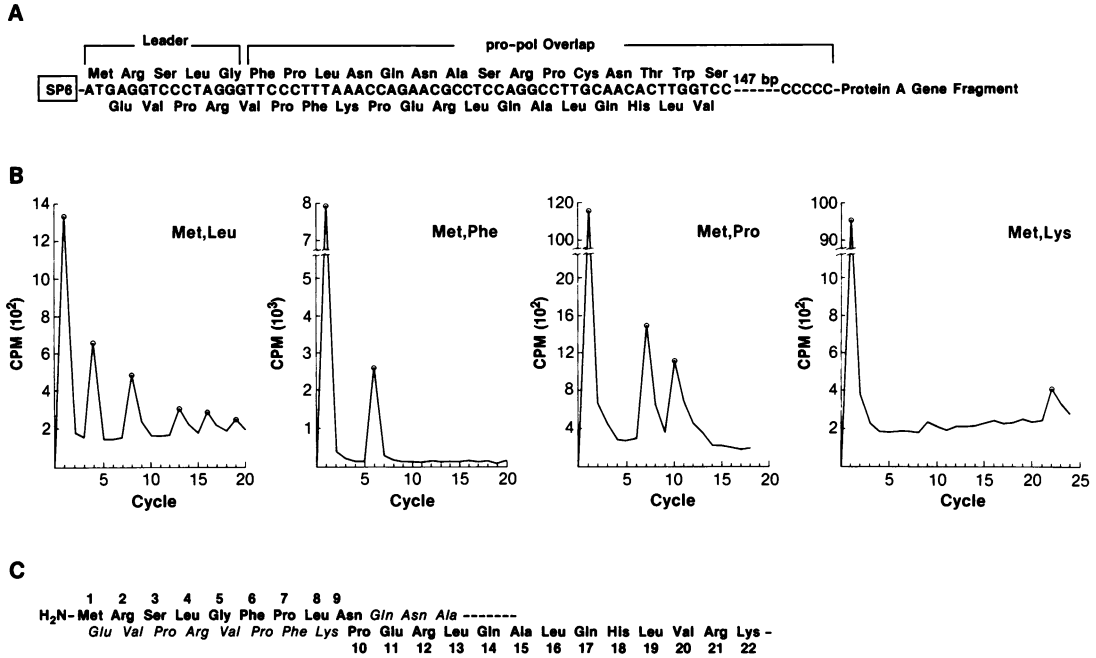


FIG. 5. Site of frameshifting in the *pro-pol* overlap. (A) A portion of plasmid pHA showing the nucleotide sequence at the 5′ end of the chimera composed of the SP6 promoter, the leader sequence, a portion of the HTLV-I *pro-pol* overlap, and the protein A gene fragment. The corresponding amino acid sequence in the *pro* frame is shown above the nucleotide sequence, and that in the *pol* frame is shown below. (B) Radioactivity profiles of Edman degration of pHA-encoded proteins synthesized in vitro and labeled with [^{35}S]methionine and another amino acid, as indicated. The radiolabeled amino acids used were [^{3}H]leucine, [^{3}H]proline, [^{3}H]phenylalanine, and [^{3}H]lysine. (C) Determined amino acid sequence of the purified transframe peptide. Amino acids 1 to 9 are derived from the *pro* frame (top row), and amino acids 10 to 22 are derived from the *pol* frame (bottom row).

3B, lane 3). These results demonstrate that the RNA sequence corresponding to 78 nt of the *Nar*I-*Stu*I fragment is essential for frameshifting to take place in the *pro-pol* overlap. In fact, this region contains the characteristic sequence UUUA, which has been shown to be the frameshift site in HIV-1 and RSV (11, 12), and is the 5′ part of the sequence UUUAAAC, which is identical to the determined heptanucleotide signal sequence in infectious bronchitis virus (2).

To define precisely the frameshift signal sequence in the *pro-pol* overlap of HTLV-I, we constructed a series of mutants containing deletions and/or base substitutions immediately upstream of the UUUA sequence (Fig. 4A). RNAs derived from plasmids pHTfsΔA and pHTfsΔB harbor heterogenous sequences between nt 2532 and 2592 obtained by replacement of the region with double-stranded synthetic oligonucleotides. The synthetic oligonucleotides retained either UUUAAAC or CCCUUUAAAC sequences (Fig. 4A). Translation of these RNAs yielded the expected Gag-Pro-Pol fusion proteins (Fig. 4, lanes 2 and 3). In contrast, deletion of the three U's upstream to the AAA run drastically reduced the synthesis of the transframe product (pHTfsΔS in Fig. 4A and B, lanes 4). To confirm the results described above, we created a 1-base substitution of A → G to produce a mutagenic construct, pHTfsΔFG, having UUUAAGC instead of UUUAAAC (Fig. 4A). As shown in lane 5 of Fig. 4B, frameshifting was severely inhibited when this mutant RNA was used in the cell-free translation system. These results demonstrate that frameshifting in the −1 direction occurs via the sequence UUUAAAC and not CCCUUUA and imply that $tRNA^{Asn}$ rather than $tRNA^{Leu}$ is involved in the frameshifting event for HTLV-I *pol* gene expression. In accordance with this conclusion, translation of the RNA derived from pHTfsCG (Fig. 4A), which harbors a 4-base substitution and has the sequence TGCAG instead of the sequence CAGAA immediately downstream of the putative signal sequence, directed the synthesis of the Gag-Pro-Pol fusion protein apparently without affecting the frameshifting efficiency observed with pHTfs (Fig. 4B, lane 6).

Amino acid sequencing of the *pro-pol* transframe protein. To identify the frameshift site in the *pro-pol* overlap, we determined a portion of the amino acid sequence of a protein produced by ribosomal frameshifting in an in vitro translation system that was similar to the one previously used by Jacks et al. (12). Plasmid pHA consists of a synthetic oligonucleotide encoding 5 amino acids of a leader sequence, including the initiator methionine, 189 nt of the HTLV-I *pro-pol* overlapping region, and an *S. aureus* protein A gene fragment (Fig. 5A; see also Materials and Methods). If ribosomal frameshifting occurs at the putative signal sequence, the site of shift from the *pro* frame to the *pol* frame should be included within amino acid positions 8 to 10. Partial amino-terminal amino acid sequences were determined through 25 cycles of Edman degradation of the pHA-encoded radiolabeled transframe protein synthesized in vitro and purified as described in Materials and Methods.

As shown in Fig. 5B, the peak of radioactive leucine encoded in the *pro* frame and that of proline encoded in the *pol* frame at positions 8 and 10, respectively, and the lack of radioactive phenylalanine and of significant amounts of lysine (*pol* frame) at positions 8 and 9, respectively, demon-

strate that the shift into the *pol* frame occurs primarily after the decoding of the asparagine codon AAC at position 9. Since radiolabeled asparagine was not used for labeling the protein, there is no direct proof of identity of the amino acid at position 9. However, the protein analysis data, together with the data from the site-directed mutagenesis studies, in which the change of the Asn codon AAC to AGC prevented the frameshift, strongly suggest that $tRNA^{Asn}$ is involved in the *pro-pol* frameshifting of HTLV-I. Only trace amounts of radioactive lysine are recovered at position 9. Thus, it appears that lysine-tRNA does not efficiently compete with asparagine-tRNA. The amino acid sequencing data, together with the results obtained by analysis of the mutants (pHTfsΔFG, pHTfsΔS, and pHTfsΔA), clearly show that the ribosomal frameshift for *pol* gene expression occurs at the slippery sequence, UUUAAAC, in which two adjacent tRNAs, the leucine-tRNA and the asparagine-tRNA, slip back simultaneously by 1 nt in the 5′ direction, and demonstrate that the shift occurs at the final 3 nt (AAC) of the slippery sequence through the mediation of $tRNA^{Asn}$.

DISCUSSION

The ORFs of HTLV-I are organized to permit the synthesis of the primary translational products of the *pro* and *pol* genes by ribosomal frameshifting in the −1 direction within the *gag-pro* and the *pro-pol* overlaps, which are 37 and 246 nt long, respectively. The Gag-Pro polyprotein, the precursor of the PR, has been shown by mutational analysis to be synthesized by frameshifting at the AAC codon of the heptanucleotide signal sequence AAAAAAC within the *gag-pro* overlap window (23). We have now defined the frameshift signal in the *pro-pol* overlap of HTLV-I and determined the exact site of the frameshift by protein sequencing. Analysis of in vitro translation products directed by mutant plasmids harboring heterogenous sequences created by the double-stranded synthetic linkers (Fig. 4) demonstrated that frameshifting in the −1 direction occurs via the heptanucleotide sequence UUUAAAC. This result suggested that the 3′ asparagine codon AAC may be the site of frameshifting. The involvement of asparagine-tRNA in frameshifting for *pol* gene expression was confirmed with a single codon substitution from AAC to AGC (serine). It should be pointed out that the conservation of three C's adjacent to the 5′ end of the frameshift signal in the *pro-pol* overlap of HTLV-I and related retroviruses results in the presence of another "signal-like" heptanucleotide sequence, CCCUUUA, in which UUUA is the consensus sequence (7). These 4 nt overlap the 5′ nucleotides of UUUAAAC, which is also the determined signal sequence in infectious bronchitis virus (3, 4) (the continuous decanucleotide sequence is CCCUUUAAAC; Fig. 1C). UUA (leucine) was shown to be the site of the frameshift in RSV (11) and HIV-1 (12). In the present study, mutant analysis demonstrated that the upstream region (CCC) of the decanucleotide sequence did not have a measurable effect on frameshifting. Moreover, sequencing of the in vitro-synthesized transframe protein provided definitive evidence that AAC is the site of the frameshift and that the C of the AAC (asparagine) codon and not the A of the UUA (leucine) codon is read twice to accomplish the frameshift (Fig. 5). Thus, it appears that in the HTLV family of retroviruses, asparagine-tRNA mediates the frameshifting at both the *gag-pro* and the *pro-pol* overlaps. In another retrovirus, MMTV, which also requires two consecutive frameshifts for the synthesis of the Gag-Pro-Pol polyprotein, asparagine-tRNA is involved in the *gag-pro* overlap (9, 13), while leucine tRNA is involved in the *pro-pol* overlap (13). It is also interesting to note that Hatfield et al. (5), who studied the tRNAs used in and near the frameshift signals in cells infected with HTLV-I and bovine leukemia virus, found that most of the asparagine-tRNA lacked the highly modified queuine (Q) base in its anticodon loop, unlike the asparagine-tRNA of uninfected cells and cells infected with other viruses, which use other shifty tRNAs for frameshifting.

Our results obtained by the analysis of mutant RNAs demonstrate that the simultaneous slippage model of −1 frameshifting in retroviruses proposed by Jacks et al. (10, 11) is also operative at the *pro-pol* overlap of HTLV-I. The results obtained by analyzing a mutant (pHTfsCG; Fig. 4) also support the proposal of Jacks et al. (12) that nucleotide changes in the sequence located immediately downstream of the frameshift signal but not affecting the putative stem-loop structure further downstream have no inhibitory effect on ribosomal frameshifting (Fig. 4).

Secondary and tertiary structure predictions of the RNA regions downstream of the putative or established ribosomal frameshifting sites (3, 36) suggested that pseudoknot formation might be required to induce optimal frameshifting. The important role of the pseudoknot structure in −1 frameshifting has been experimentally established in studies of infectious bronchitis virus (3) and MMTV (4). In addition to the experiments that we describe in this report, we have already obtained evidence that complete disruption of the putative pseudoknot (including the proximal and second stem-loop structures) prevents *pro-pol* frameshifting in HTLV-I. However, further experiments with quantitative assays will be required to obtain insights into how the secondary and tertiary RNA structures downstream of the signal sequence influence the efficiency of frameshifting in the *gag-pro* and *pro-pol* overlaps of HTLV-I.

ACKNOWLEDGMENTS

We thank Young D. Kim for N-terminal amino acid sequence analysis, Alan Rein for useful discussions, and Dolph Hatfield and Judith Levin for critical reading of the manuscript. We are also grateful to Cheri Rhoderick for typing the manuscript and Richard Frederickson for graphics assistance.

This research was sponsored in part by the National Cancer Institute under contract NO1-CO-74101 with ABL.

REFERENCES

1. **Birnboim, H. C., and J. Doly.** 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. **7**:1513–1523.
2. **Brierley, I., M. E. G. Boursnell, M. M. Binns, B. Bilimoria, V. C. Block, T. D. K. Brown, and S. C. Inglis.** 1987. An efficient ribosomal frameshifting signal in the polymerase-encoding region of the coronavirus IBV. EMBO J. **6**:3779–3785.
3. **Brierley, I., P. Digard, and S. C. Inglis.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57**:537–547.
4. **Chamorro, M., N. Parkin, and H. E. Varmus.** 1992. An RNA pseudoknot and an optimal heptameric shift site are required for highly efficient ribosomal frameshifting on a retroviral messenger RNA. Proc. Natl. Acad. Sci. USA **89**:713–717.
5. **Hatfield, D., Y.-X. Feng, B. J. Lee, A. Rein, J. G. Levin, and S. Oroszlan.** 1989. Chromatographic analysis of the aminoacyl-tRNAs which are required for translation of codons at and around the ribosomal frameshift sites of HIV, HTLV-1 and BLV. Virology **173**:736–742.
6. **Hatfield, D., and S. Oroszlan.** 1990. The where, what and how of ribosomal frameshifting in retroviral protein synthesis. Trends Biochem. Sci. **15**:186–190.
7. **Hatfield, D. L., J. G. Levin, A. Rein, and S. Oroszlan.** 1992.

Translational suppression in retroviral gene expression. Adv. Virus Res. **41**:193–239.
8. **Hinuma, Y., K. Nagata, M. Hanaoka, M. Nakai, T. Nakai, K. Kinoshita, S. Shiragawa, and I. Miyoshi.** 1981. Adult T-cell leukemia: antigen in an ATL cell line and detection of antibodies to the antigen in human sera. Proc. Natl. Acad. Sci. USA **78**:6476–6480.
9. **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.
10. **Jacks, T.** 1990. Translational suppression in gene expression in retroviruses and retrotransposons. Curr. Top. Microbiol. Immunol. **157**:93–124.
11. **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.
12. **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.
13. **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.
14. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230**:1237–1242.
15. **Krieg, P. A., and D. A. Melton.** 1984. Functional messenger RNAs are produced by SP6 *in vitro* transcription of cloned cDNAs. Nucleic Acids Res. **12**:7057–7070.
16. **Leis, J., D. Baltimore, J. M. Bishop, J. Coffin, E. Fleissner, S. P. Goff, S. Oroszlan, H. Robinson, A. M. Skalka, H. M. Temin, and V. Vogt.** 1988. Standardized and simplified nomenclature for proteins common to all retroviruses. J. Virol. **62**:1808–1809.
17. **Mador, N., A. Panet, and A. Honigman.** 1989. Translation of *gag*, *pro*, and *pol* gene products of human T-cell leukemia virus type 2. J. Virol. **63**:2400–2404.
18. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
19. **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.
20. **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.
21. **Mori, K., H. Sabe, H. Siomi, T. Iino, A. Tanaka, K. Takeuchi, K. Hirayoshi, and M. Hatanaka.** 1987. Expression of a provirus of human T-cell leukaemia virus type I by DNA transfection. J. Gen. Virol. **68**:499–506.
22. **Nam, S. H., and M. Hatanaka.** 1986. Identification of a protease gene of human T-cell leukemia virus type I (HTLV-I) and its structural comparison. Biochem. Biophys. Res. Commun. **139**: 129–135.
23. **Nam, S. H., M. Kidokoro, H. Shida, and M. Hatanaka.** 1988. Processing of Gag precursor polyprotein of human T-cell leukemia virus type I by virus-encoded protease. J. Virol. **62**:3718–3728.
24. **Poiesz, B. J., F. W. Ruscetti, A. F. Gazdar, P. A. Bunn, J. D. Minna, and R. C. Gallo.** 1980. Detection and isolation of type C retrovirus particles from fresh and cultured lymphocytes of a patient with cutaneous T-cell lymphoma. Proc. Natl. Acad. Sci. USA **77**:7415–7419.
25. **Poiesz, B. J., F. W. Ruscetti, M. S. Reitz, V. S. Kalyanaraman, and R. C. Gallo.** 1981. Isolation of a new type C retrovirus (HTLV) in primary uncultured cells of a patient with Sezary T-cell leukemia. Nature (London) **294**:268–271.
26. **Power, M. D., P. A. Marx, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1986. Nucleotide sequence of SRV-1, a type D simian acquired immune deficiency syndrome retrovirus. Science **231**:1567–1572.
27. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
28. **Rice, N. R., R. M. Stephens, A. Burny, and R. V. Gilden.** 1985. The *gag* and *pol* genes of a bovine leukemia virus: nucleotide sequence and analysis. Virology **142**:357–377.
29. **Sagata, N., T. Yasunaga, J. Tsuzuku-Kawamura, K. Ohishi, Y. Ogawa, and Y. Ikawa.** 1985. Complete nucleotide sequence of the genome of bovine leukemia virus: its evolutionary relationship to other retroviruses. Proc. Natl. Acad. Sci. USA **82**:677–681.
30. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
31. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32**:853–869.
32. **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.
33. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
34. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45**:375–385.
35. **Taylor, J. W., J. Ott, and F. Eckstein.** 1985. The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA. Nucleic Acids Res. **13**:8765–8785.
36. **ten Dam, E. B., C. W. A. Pleij, and L. Bosch.** 1990. RNA pseudoknot: translational frameshifting and readthrough on viral RNAs. Virus Genes **4**:121–136.
37. **Uhlen, M., B. Guss, B. Nilsson, S. Gatenbeck, L. Philipsons, and M. Lindberg.** 1984. Complete sequence of the staphylococcal gene encoding protein A. A gene evolved through multiple duplications. J. Biol. Chem. **258**:1685–1702.
38. **Yoshida, M., I. Miyoshi, and Y. Hinuma.** 1982. Isolation and characterization of retrovirus from cell lines of human adult T-cell leukemia and its implication in the disease. Proc. Natl. Acad. Sci. USA **79**:2031–2035.
39. **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.
40. **Yoshinaka, Y., and S. Oroszlan.** 1985. Bovine leukemia virus post-envelope gene coded protein: evidence for expression in natural infection. Biochem. Biophys. Res. Commun. **131**:347–354.