JOURNAL OF VIROLOGY, Mar. 1992, p. 1737–1745
0022-538X/92/031737-09$02.00/0

Vol. 66, No. 3

Complex Splicing in the Human T-Cell Leukemia Virus (HTLV) Family of Retroviruses: Novel mRNAs and Proteins Produced by HTLV Type I

VINCENZO CIMINALE,[1] GEORGE N. PAVLAKIS,[1] DAVID DERSE,[2] CHRISTIAN P. CUNNINGHAM,[3] AND BARBARA K. FELBER[3]*

Human Retrovirus Section[1] and Human Retrovirus Pathogenesis Group,[3] ABL-Basic Research Program, Laboratory of Viral Carcinogenesis,[2] National Cancer Institute, Frederick Cancer Research and Development Center, Frederick, Maryland 21702-1201

Received 14 August 1991/Accepted 13 December 1991

Novel cytoplasmic mRNA species produced by human T-cell leukemia virus type I (HTLV-I) were cloned by using the polymerase chain reaction technique. Five novel 3′ splice sites located in the X region and upstream of the *env* gene were identified. Splicing to the 3′ splice sites in the X region generates mRNAs that express two previously unidentified viral proteins, named Rof and Tof. Tof accumulates in the nucleoli of transfected cells. The other viruses of the HTLV family, such as HTLV-II and bovine leukemia virus, also have a complex splicing pattern and are capable of producing additional proteins encoded in the X region. These results suggest that HTLV-I and other members of the HTLV family produce novel proteins, which may contribute to the biological properties of these viruses.

The human T-cell leukemia virus (HTLV) family includes HTLV types I and II (HTLV-I and HTLV-II), simian T-cell leukemia virus type I, and bovine leukemia virus (BLV). These viruses produce essential proteins, in addition to Gag, Pol, and Env, that regulate viral expression. This characteristic is shared by the lentiviruses, including human immunodeficiency virus type 1 (HIV-1). Therefore, it has been proposed that the lentiviruses and the HTLVs should be classified together in a group designated complex retroviruses (8, 30, 33).

The two regulatory proteins necessary for expression of the HTLVs are Tax and Rex. Tax is a transcriptional activator of the viral promoter (6, 10, 16, 34, 41, 43). Rex affects posttranscriptional regulatory steps by promoting transport of the unspliced and singly spliced mRNA from the nucleus to the cytoplasm and by promoting expression of the Gag, Pol, and Env proteins (11, 21, 22, 24, 40). Tax and Rex are expressed from two overlapping reading frames located in the distal part of the X region of the virus by a bicistronic viral mRNA, consisting of three exons (Fig. 1B) (31, 39). The X region lies between the *env* terminator and the 3′ long terminal repeat (LTR) and contains four open reading frames (ORFs), initially named x-I, x-II, x-III, and x-IV (Fig. 1). The x-IV ORF encodes the second exon of Tax, while the x-III ORF encodes the second exon of Rex. The proximal part of the X region in HTLV-I consists of 655 nucleotides (nt) containing ORFs x-I and x-II, which are not known to be expressed. This organization is common to all retroviruses of the HTLV family since simian T-cell leukemia virus type I, HTLV-II, and BLV have a similar arrangement (see Fig. 5). We investigated whether additional mRNAs are produced by HTLV-I, HTLV-II, and BLV via alternative splicing in the X region to express the additional ORFs.

* Corresponding author.

MATERIALS AND METHODS

Recombinant constructs. CS-HTLV-I DNA was cloned from cell line CS-1, obtained by cocultivation of cord blood lymphocytes with lethally irradiated HS-1 cells. The latter is an HTLV-I-infected B-lymphocyte cell line derived from the peripheral blood of a 20-year-old female with adult T-cell leukemia (26, 27). CS-1 cells harbor three intact HTLV-I proviruses. DNA from CS-1 cells was digested to completion with *Eco*RI and cloned into the bacteriophage lambda DASH vector (Stratagene). One of the cloned HTLV-I proviruses was then subcloned into the plasmid vector pUC19 such that cellular flanking sequences were removed. The HTLV-I plasmid clone, termed pCS-HTLV-I, lacks 30 bp from the left end of the 5′ LTR and approximately 60 bp from the distal end of the 3′ LTR. These missing sequences do not alter expression of the transfected virus and are regenerated after one round of virus replication. Eucaryotic expression plasmids were constructed by inserting the polymerase chain reaction (PCR)-amplified *rof* and *tof* cDNAs into an expression vector which contains the HIV-1 LTR promoter, the Rev-responsive element, and the polyadenylation signal. All other constructs used have been described in detail elsewhere. pL10CAT contains the HTLV-I LTR promoter linked to the chloramphenicol acetyltransferase indicator gene (32), pLcXL expresses the *tax* cDNA from the HTLV-I LTR promoter, pCgag-RXRE (5) contains the HIV-1 *gag* gene linked to Rex-responsive element of HTLV-I (RXRE), and pL3rex contains the *rex* gene linked to the HIV-1 LTR promoter (44).

Transfections and protein analysis. HLtat cells (36), which constitutively produced the HIV-1 Tat, were transfected as described previously (14, 18), and two days later, the cells were harvested and analyzed for RNA or protein production. Chloramphenicol acetyltransferase expression was quantitated as described previously (17). HIV-1 $p24^{gag}$ expression was analyzed by using an enzyme-linked immunosorbent antigen-capture assay (NEN). Indirect immunofluorescence was performed as described elsewhere (16).

RNA isolation, cDNA synthesis, PCR amplification, and

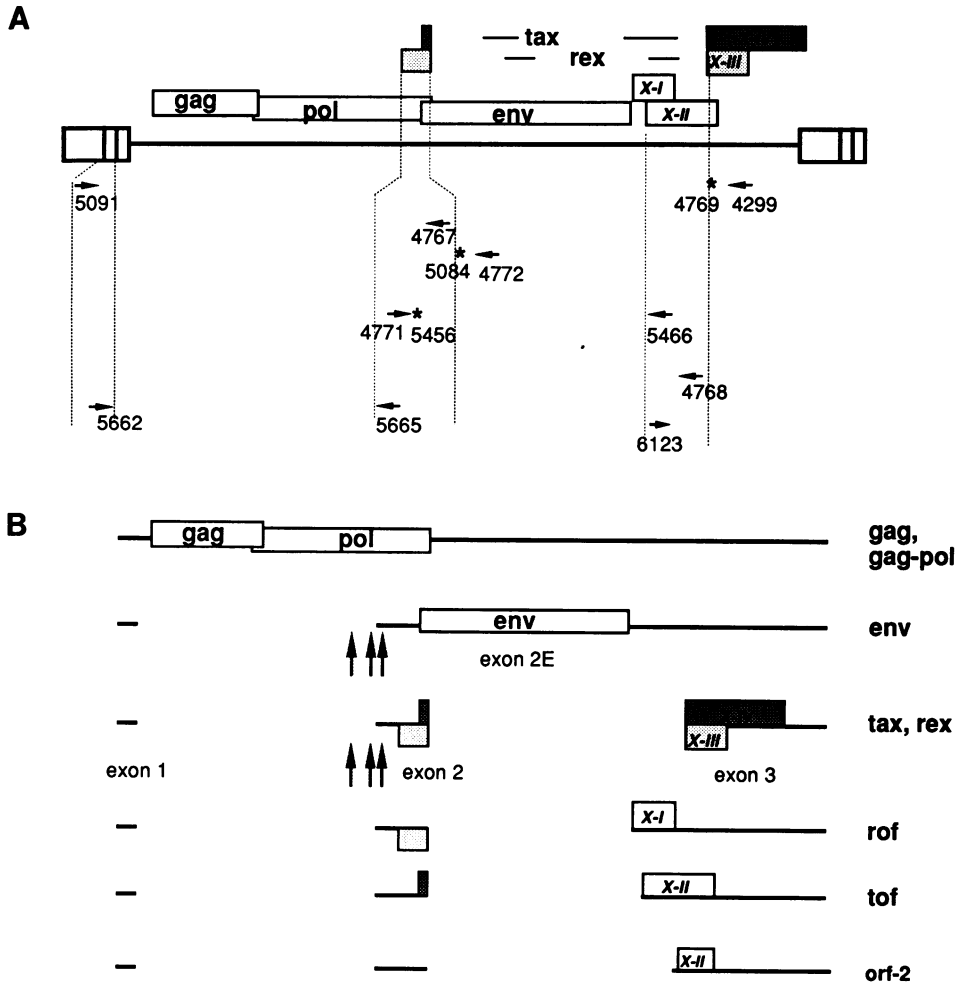

FIG. 1. (A) Organization of the HTLV-I genome. The viral ORFs are indicated. The positions of the primers used for amplification are indicated by arrows; the oligonucleotides used as hybridization probes are indicated by asterisks. The numbering follows the sequence of Seiki et al. (39) as reported previously (29) in which +1 is the first nucleotide of the viral mRNA at the beginning of the R region. Cytoplasmic RNA was reverse transcribed, and the resulting cDNAs were amplified with the PCR technique by using the indicated primer pairs. (B) The structures and coding potential of the different HTLV-I mRNAs are shown. Arrows indicate the three alternative splice sites for exon 2 used in the *env* and the doubly spliced mRNAs.

cloning. Cytoplasmic RNA (9, 15) (0.1 to 0.4 μg) was reverse transcribed in the presence of 500 ng of a mix of random hexamers [pd(N)$_6$; Pharmacia] for 2 h at 45°C, and the resulting cDNAs were amplified by the PCR in the same tube as previously described (36). The resulting cDNAs were cloned into the Bluescript KS$^-$ vector (Stratagene). For sequence analysis of the cloned cDNAs, single-stranded DNA was prepared and sequenced according to standard protocols.

The names, sequences, and exact locations in the HTLV-I genome of the oligonucleotides used in this work are as follows: 5091, 5′-GGGCTCGCATCTCTCCTTCACGCG-3′ (nt 1 to 24); 4771, 5′-CTTCCTGGTCTTAATAGCCGCC-3′ (nt 4641 to 4662); 4772, 5′-CCGAAGATGAGGGGGCAGAACTG GAAG-3′ (nt 4886 to 4860); 5084, 5′-GGAAGAATAA A A T CAAAGTGGCGAGAAACTTACCC-3′ (nt 4864 to 4830); 4767, 5′-GGGATCGGCGGGGCCTCCGACGGG-3′ (nt 4805 to 4782); 5466, 5′-GCTGAGAAGGCGAAACAGC-3′ (nt 6502 to 6484); 4768, 5′-CTGATAATAAGCATGGTTAACTTTG-3′ (nt 6949 to 6925); 4299, 5′-AACACGTAGACTGGGTATCC-3′ (nt 7004 to 6985); 5662, 5′-CTGCGTCCGCCGTCTAG-3′ (nt 103 to 119); 5665, 5′-GGCGGCTATTAAGACCAGGAAG-3′ (nt 4662 to 4641); 5456, 5′-CCAGTGGAAAGGACCAC AGG-3′ (nt 4661 to 4680); 4769, 5′-GACTCTGTCCAAAC-CCTGGGAAGTGGG-3′ (nt 6976 to 6950); and 6123, 5′-CCAACACCATGGCACTATGCTGTTTCGCCTTCTCAGC C-3′ (nt 4820 to 4831 and nt 6478 to 6503). Sequence analysis of the CS-HTLV-I isolate revealed the following differences with the published sequence (39): nt 129, A to G; nt 4819, T to A; nt 6632, G to A; nt 6633, G to A; nt 6668, T to C; nt 6744, G to A; nt 6752, T to C; nt 6806, T to C; nt 6892, A to G; nt 6901, A to G; nt 6918, G to A; nt 7026, C to T; and nt 8406, A to G.

The names, sequences, and positions of the HTLV-II specific oligonucleotides are as follows: 6004, 5′-CCCCCCGGTGGATCC-3′ (nt 5081 to 5095); 6005, 5′-GAT TCCTGAAAAAAGCTGCATGC-3′ (nt 5102 to 5124); 6012, 5′-GGAAGAAAGGCATCAGCAGT-3′ (nt 6911 to 6892); and 6013, 5′-AGCGGGAGCGGCGCAGA-3′ (nt 6969 to

6953). The numbering follows the published sequence (42), whereby +1 is the first nucleotide of U3. Sequence analysis of the HTLV-II X region revealed the following differences with the published sequence: nt 6744, A to G; nt 6751, T to A; nt 6752, deletion of T; nt 6824, insertion of C; and nt 7223, T to C.

The names, sequences, and positions of the BLV specific oligonucleotides are 6101, 5′-AGGCGCTCTCCTGGCTAC TG-3′ (nt 4758 to 4777); 6100, 5′-GAAAGCCTTCAAATGCCT AAAGA-3′ (nt 4809 to 4831); and 6102, 5′-CAACCTTCCT CAGGGTGCGA-3′ (nt 6863 to 6844). The numbering follows the published sequence (35), whereby +1 is the first nucleotide of U3. Sequence analysis of the pBLV913 isolate revealed the following difference with the published sequence (35): nt 4830, G to A.

For Northern (RNA) blot analysis, HLtat cells were transfected with CS-HTLV-I DNA, total RNA was prepared by the RNAzol method (Cinna/Biotecx), and polyadenylated mRNAs were selected twice by oligo(dT) chromatography according to the manufacturer's recommended protocol (Pharmacia) and subjected to Northern analysis as described previously (19). HTLV-I RNAs were detected by hybridization to a ^{32}P-labeled DNA probe generated by PCR amplification of CS-HTLV-I DNA by using primer pair 6123 and 4768 that is specific for *rof* and *tof* mRNAs. Radioactivity on resulting blots was quantitated with the AMBIS radioanalytic imaging system.

In vitro transcription and translation. Amplified cDNAs were cloned into the *Xho*I site of the Bluescript KS$^-$ vector (Stratagene). The resulting plasmids were linearized with *Hin*dIII, and capped run-off transcripts were synthesized with T3 RNA polymerase (Stratagene) according to the recommendations of the manufacturer. In vitro translation was performed in a nuclease-treated rabbit reticulocyte lysate (Promega) in the presence of 25 μCi of [^{3}H]leucine. The produced proteins were analyzed on SDS–15% acrylamide gels (25). Immunoprecipitations were performed as described previously (14).

RESULTS

Novel mRNAs expressing the x-I and x-II ORFs of HTLV-I. A molecular clone of HTLV-I generated from lymphocytes of a patient with adult T-cell leukemia and named CS-HTLV-I was used to study the production of viral mRNAs. Upon introduction into mammalian cells, CS-HTLV-I produced all viral proteins, including Tax and Rex. Electron microscopy studies demonstrated that mature virus particles are produced from this molecular clone. We reasoned that this apparently intact molecular clone would produce all viral mRNAs after transfection of proviral DNA into human cells and would provide an ideal model for studying expression of HTLV-I. Cell lines persistently infected with HTLV-I often contain defective proviruses producing additional transcripts, which may complicate the interpretation of the results.

First, we investigated whether HTLV-I expressed mRNAs encoding the x-I and x-II ORFs. HeLa cells were transfected with CS-HTLV-I, and 2 days later, cytoplasmic RNA was isolated and reverse transcribed with a mix of random hexanucleotide primers. The resulting cDNAs were amplified by PCR with specific oligonucleotide primers, as indicated in Fig. 1A. The primer pairs were designed to specifically detect mRNAs that splice from exon 2 to the x-I and x-II ORFs. The resulting amplification products were separated on acrylamide gels (Fig. 2A), blotted onto nitrocellulose filters, and hybridized with a ^{32}P-labeled internal oligonucleotide (5456) as probe. The results of this analysis confirmed that the three bands produced by PCR amplification were HTLV-I specific (lanes 4 and 5). The cDNAs were

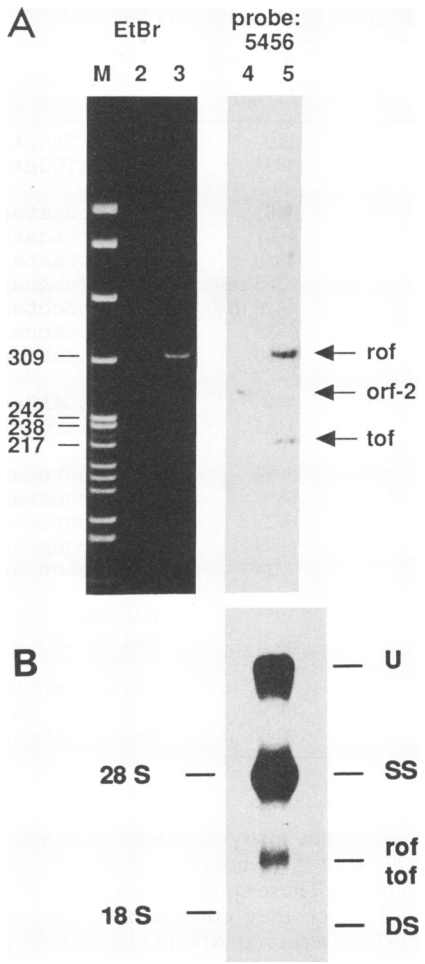

FIG. 2. (A) Southern blot of PCR-amplified cDNAs containing portions of the x-I and x-II regions. Primer pairs 4771 and 5466 and 4771 and 4768 were designed to specifically detect mRNAs that splice from exon 2 to the x-I and x-II ORFs. Portions (10 μl) of each PCR amplification were analyzed on 5% nondenaturing polyacrylamide gels. DNA was visualized by UV fluorescence after being stained with ethidium bromide (EtBr, lanes 2 and 3), transferred to nylon membranes, and hybridized as previously described (36) to probe 5456 (lanes 4 and 5). The assignment of the amplified bands to different mRNAs is shown on the right. The size markers are indicated to the left. Lanes: 1, *Msp*I-digested pBR322; 2 and 4, amplification with primer pair 4771 and 4768; 3 and 5, amplification with primer pair 4771 and 5466. (B) Northern blot analysis of poly(A)$^+$ mRNA from cells transfected by CS-HTLV-I. Approximately 4 μg of poly(A)$^+$-selected mRNA from HLtat cells transfected by pCS-HTLV-I DNA was analyzed on a denaturing gel and blotted. HTLV-I RNAs were detected by hybridization to a ^{32}P-labeled probe specific for unspliced, singly spliced, and *rof* and *tof* RNAs, generated by PCR amplification with primers 6123 and 4768. A duplicate blot was hybridized with a probe able to detect all HTLV-I mRNAs (not shown). The positions of the three characterized classes of HTLV mRNAs (unspliced [U], singly spliced [SS], and doubly spliced [DS]) are indicated on the right. The positions of 28S and 18S rRNAs are indicated on the left. Radioactivity on resulting blots was quantitated with the AMBIS radioanalytic imaging system.

TABLE 1. Splice sites used in HTLV-I, HTLV-II, and BLV

Virus	Splice site	Nucleotide position no.	Base sequence
HTLV-I	5′	119	GCCGTCTAG**gtaagtttag**
		4831	ACACCATGG**gtaagtttct**
	3′	4501	**ctctatccatag**CCCTAT
		4641	**tattatttcaag**CTTCCT
		4658	**tggtcttaatag**CCGCCA
		6383 (*rof*)	**cttctcctgcag**CAACTT
		6478 (*tof*)	**tttctttcctag**CACTAT
		6875	**ttcctccaccag**CAGGTC
		6950	**gcttattatcag**CCCACT
HTLV-II	5′	449	TCCTCCCAAG**gtaagtctcc**
		5183	AACACCATGG**gtaatgtttt**
	3′	5044	**caatccctccag**GAAGCG
		6807	**ttcctcccatag**GACCTT
		6827	**ttccccttccag**GAAATC
		6944	**cttttcccgcag**GCGCTC
		7214	**gtctcctctcag**CCCATT
BLV	5′	305	CAGCGGTCAG**gtaaggca**
		4871	CATCAGATGG**gtaagtct**
	3′	4649	**ctgtcatttcag**AGGGCG
		6824	**caccctttctag**AGATAC
		7247	**ttgtcttttaag**CAAGTG

cloned, and sequence analysis of several clones resulted in the identification of three novel 3′ splice sites (3′ SS) in the X region (Table 1). These splice sites agree with the consensus for 3′ SS, since they contain pyrimidine-rich regions upstream of the conserved AG dinucleotide at the splicing site. Primer pair 4771 and 5466 specifically amplified cDNA fragments from mRNAs in which exon 2 was spliced to either of two 3′ SS located at nt 6383 or nt 6478. Primer pair 4771 and 4768 specifically amplified cDNA fragments from mRNAs containing exon 2 spliced to a 3′ SS at nt 6875. These splicing events generate mRNAs that produce new proteins. Splicing to nt 6383 generates an ORF starting at the Rex AUG and continuing in the x-I ORF. This ORF consists of 152 amino acids and was named Rex-xI ORF, or Rof. Splicing to nt 6478 generates an ORF consisting of the first exon of Tax linked to the x-II ORF. This ORF was named Tax-xII ORF, or Tof, and consists of 241 amino acids. Splicing to the 3′ SS at nt 6875 generates an mRNA, which may encode a smaller protein of 87 amino acids starting at an internal AUG in x-II ORF at nt 6936. RNAs containing the first exon of HTLV-I spliced directly to these novel 3′ SS were not detected by amplification with primer pair 5091 and 5466 or 5091 and 4768. To determine whether any further splicing occurred in these mRNAs 3′ to the x-I or x-II region, cDNAs were PCR amplified by using primer 6123 located in the x-I/x-II region and primer 4299 located in the last exon. Only the expected band of 538 nt corresponding to the unspliced mRNA was detected (data not shown), suggesting that these mRNAs do not undergo further splicing.

To verify the presence of these mRNAs in HTLV-I-infected cell lines, cytoplasmic RNA was prepared from cell lines chronically infected with HTLV-I (MT-2 and C91-PL), reverse transcribed, and PCR amplified as described above. The same bands were detected in these experiments (data not shown), demonstrating that the three novel 3′ SS in the x-I and x-II regions are utilized in both HTLV-I-transfected and -infected cells.

Northern blot analysis of poly(A)$^+$-selected mRNA from HeLa cells transfected with CS-HTLV-I revealed the presence of additional mRNAs when probed with an x-I- and x-II-specific probe (Fig. 2B). These mRNAs were of the expected size predicted for *rof* and *tof* mRNAs. To detect the low levels of the *rof* and *tof* mRNAs, approximately 4 μg of poly(A)$^+$ mRNA selected by being passed twice through an oligo(dT) column (Pharmacia) was loaded on the lane. This indicated that these mRNAs are expressed at low levels in transiently transfected cells. On the basis of these experiments, we estimated that *rof* and *tof* mRNAs are approximately 20-fold less abundant than the *env* mRNA. It is possible that *tof* and *rof* mRNAs are expressed at higher levels at certain times after infection or in different cell types. It has been shown previously that the levels of expression of different mRNAs produced by HIV-1 vary greatly during the viral life cycle (1, 12, 14, 23). In conclusion, the combination of PCR amplification and Northern blot analysis demonstrated the presence of additional HTLV-I mRNAs.

Additional alternative splice acceptors found for exon 2. To identify additional splice sites between exons 1 and 2 in HTLV-I, two primer pairs (5091 and 5665 and 5662 and 5665) were chosen for amplification of the CS-HTLV-I cDNAs. Sequence analysis of the cloned cDNAs revealed the presence of two novel alternative 3′ SS for exon 2 (Table 1 and Fig. 1B). These 3′ SS are located 140 nt 5′ and 17 nt 3′ to the previously described 3′ SS at nt 4641, which generates the known singly and doubly spliced mRNAs. Splicing to these sites generates alternative exons to the known exon 2. These alternative exons do not affect the coding capacity of the mRNAs, since the same proteins are predicted to be produced by the different mRNAs. The same three alternative 3′ SS were detected by using a primer pair (5091 and 4772) specific for singly spliced mRNAs. These data reveal these three alternative splice sites can be used to generate the *env* and all the doubly-spliced mRNAs.

It is noteworthy that all mRNAs for Tax, Tof, and Env contain Rex as the first ORF. Using the primer pair 5091 and 4767, we investigated whether additional mRNAs were generated by splicing downstream of the Rex AUG in exon 2 (data not shown). Such mRNAs would contain the Tax/Env AUG as the first AUG. These experiments failed to detect any such mRNAs. Therefore, proteins beginning with the Tax/Env AUG are exclusively produced from bicistronic mRNAs (31).

Expression of *rof* and *tof* cDNAs in vitro. To identify the protein products generated by the *rof* and *tof* cDNAs, these cDNAs were cloned into the Bluescript KS$^-$ vector (Stratagene) and transcribed in vitro using the T3 RNA polymerase promoter. The produced mRNAs were translated by using the rabbit reticulocyte lysate, and the resulting proteins were analyzed on SDS-polyacrylamide gels (Fig. 3). The *rof* cDNA was predicted to produce a protein of 17 kDa, which migrated at 27 kDa on 15% denaturing polyacrylamide gels (Fig. 3B, lane 1). It is noteworthy that Tat, Rex, and Rev, which contain charged regions with high arginine and lysine content, also migrate slower than expected in polyacrylamide gels. Tof, a protein of a predicted size of 26.5 kDa, migrated at 30 kDa (Fig. 3C).

We generated rabbit antisera against peptides corresponding to different portions of Rof and Tof proteins. The antiserum generated against the carboxy terminus of Tof (amino acids 225 to 241) was able to specifically immunopre-

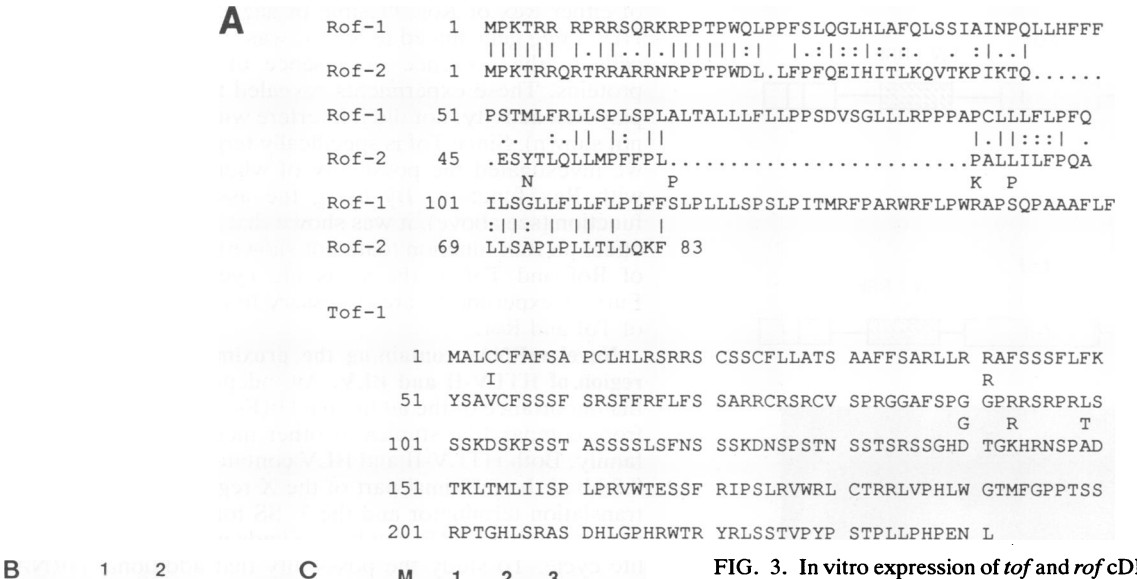

FIG. 3. In vitro expression of *tof* and *rof* cDNAs. (A) Amino acid sequences of Rof and Tof of HTLV-I and comparison of HTLV-I Rof (Rof-1) to HTLV-II Rof (Rof-2). The predicted sequences of Rof-1 and Tof-1 are according to reference 39. The differences with our clone (CS-HTLV-I) are indicated at the top. The Rof-2 sequence has been corrected to reflect the identified differences (see Materials and Methods). (B and C) In vitro translation of *rof* (B) and *tof* (C) mRNAs. Lanes (panel B): 1, in vitro translation of *rof* mRNA; 2, molecular weight markers. Lanes (panel C): M, molecular weight markers; 1, immunoprecipitation of in vitro synthesized Tof with preimmune rabbit serum; 2, in vitro translation of *tof* mRNA; 3, immunoprecipitation of in vitro-synthesized Tof protein with antiserum generated against the carboxy terminus of Tof.

cipitate the in vitro-produced Tof protein (Fig. 3C, lane 3). In contrast, the antiserum generated against a carboxy-terminal Rof peptide (amino acids 138 to 152) did not recognize in vitro-synthesized Rof. The failure to generate Rof-reactive antiserum might be caused by the hydrophobic character of this protein.

Tof is a nuclear protein. To examine the properties of Tof and Rof in vivo, the cDNAs were cloned into a eukaryotic expression vector (Fig. 4A). This vector contains the HIV-1 LTR promoter, the Rev-responsive element, and the polyadenylation site and offers two advantages: a high level of expression in the presence of Tat in human cells and the possibility to stimulate expression of defective mRNAs in the presence of Rev or Rex, as previously described (13, 15, 19). Rex is necessary for the expression of HTLV-I *gag* and *env* mRNAs, and it could also be needed for the expression of *rof* and *tof* mRNAs.

A HeLa cell line (HLtat) (36) that constitutively expresses the HIV-1 Tat protein was transfected with the *tof* expression plasmid. One day later, the cells were fixed and analyzed by indirect immunofluorescence. Figure 4B demonstrates that Tof accumulates in the nucleoli of transfected cells. No difference in Tof expression was detected in the presence or absence of the Rev protein, indicating that expression of Tof was independent of Rev and Rex (data not shown).

Rof and Tof do not affect the function of Tax and Rex. The predicted amino acid sequences of the Rof and Tof proteins are shown in Fig. 3A. Rof contains the arginine-rich region important for the binding of Rex (2) to RXRE. The rest of the protein is predicted to be highly hydrophobic and has high proportions of leucine, proline, and phenylalanine. The leucine-rich regions show homology to hydrophobic domains of several known membrane proteins.

Tof contains regions rich in serine and threonine. Computer searches of the protein data bases revealed similarities between Tof and serine-rich proteins, including several transcriptional activators such as oct-1, oct-2, pit-1, engrailed, and POU-M1. The serine/threonine-rich regions have been shown to be necessary activation domains in these proteins (20, 45, 46). Since this analysis raises the possibility that Tof may function as an activator of transcription, we investigated whether it could affect expression of the viral LTR promoter. Plasmid pL10CAT, which contains the HTLV-I LTR promoter linked to the chloramphenicol acetyltransferase indicator gene, was transfected into HLtat cells together with the *tof* expression plasmid in the presence or absence of pLcXL, which expresses HTLV-I Tax. Two days later, the cells were harvested and analyzed for chloramphenicol acetyltransferase expression (data not shown). These experiments showed that Tof was not able to activate expression from the LTR promoter and it did not affect the extent of transactivation by Tax protein.

The lack of monospecific antibodies against Rof did not allow the detection of Rof in human cells. Since Rof contains the first exon of Rex, which encodes the RNA-binding domain, we also tested whether Rof has Rex function by measuring the activation of *gag* expression in the presence

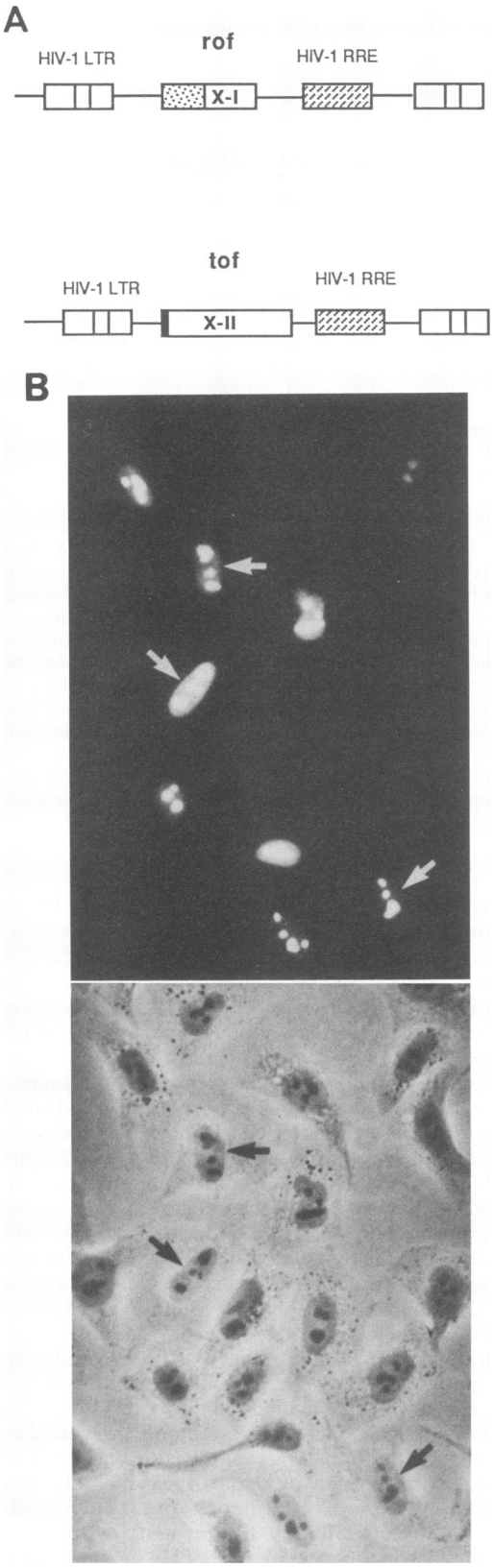

FIG. 4. Tof is a nucleolar protein. (A) Eukaryotic expression vectors for *rof* and *tof* cDNAs. The HTLV-I *tof*- and *rof*-coding regions were cloned after the HIV-1 LTR promoter in the expression plasmid, which also provides the HIV-1 Rev-responsive element (RRE) and polyadenylation signal. (B) Indirect immunofluorescence of Tof. The *tof* expression plasmid was transfected into HLtat cells, which were fixed 1 day later. Indirect immunofluorescence was performed by using the Tof antiserum generated against a carboxy-terminal peptide of Tof. Mock-transfected cells (not shown) were scored negative in this assay. Results of phase-contrast microscopy (bottom) and immunofluorescence (top) are shown. Arrows indicate some of the positive cells.

of either Rex or Rof. Plasmid pCgag-RXRE containing the HIV-1 *gag* gene linked to RXRE was transfected into HLtat cells in the presence or absence of Rex, Rof, or both proteins. These experiments revealed that Rof did not display Rex activity, nor did it interfere with Rex function (data not shown). Since Tof is specifically targeted to the nucleoli, we investigated the possibility of whether it may interfere with Rex function. By using the assay to test for Rex function (see above), it was shown that Tof did not have any effect on Rex function (data not shown). Therefore, the role of Rof and Tof in the virus life cycle remains unclear. Further experiments are necessary to evaluate the function of Tof and Rof.

Novel mRNAs containing the proximal portion of the X region of HTLV-II and BLV. An independent argument for the importance of the additional ORFs in the X region arises from comparative studies of other members of the HTLV family. Both HTLV-II and BLV contain a region larger than 0.5 kb at the proximal part of the X region between the *env* translation terminator and the 3′ SS for Tax and Rex. This region contains ORFs but has no known function in the viral life cycle. To study the possibility that additional mRNAs expressing this part of the genome were produced by the different viruses, we followed a strategy similar to the one outlined above for HTLV-I. We isolated cytoplasmic RNA from either HTLV-II-infected cells (line MO) (6a) or cells transfected with a molecular clone of BLV (pBLV913 [13]). Reverse transcriptase-PCR analysis (primer pair 6004 and 6012 and primer pair 6004 and 6013 for HTLV-II; primer pair 6101 and 6102 for BLV) revealed the presence of novel cDNAs. Sequence analysis of cloned PCR products revealed the presence of additional mRNAs containing the proximal part of X region. HTLV-II possesses three additional 3′ SS at nt 6807, 6827, and 6944, whereas BLV possesses one additional 3′ SS at nt 6824 (Table 1 and Fig. 5). With HTLV-II, splicing to nt 6807 generates two ORFs of 83 amino acids each, starting at the Rex and the Tax AUGs, respectively. Splicing to nt 6827 generates similar proteins containing seven amino acid in-frame deletions. In BLV, splicing of exon 2 to nt 6824 generates a 94-amino-acid protein initiating at the Tax AUG and including the x-I region. Interestingly, the Rof proteins of HTLV-I and HTLV-II show significant homology (63% similarity, Fig. 3A). Alignment of the x-I region of BLV (Fig. 5) also indicates some homology (38% similarity), suggesting that this protein is well conserved among the members of the HTLV family.

Analysis of sera from HTLV-I-positive individuals. To study the expression of Rof and Tof proteins in HTLV-I-infected people, we analyzed 28 serum samples obtained from different groups of HTLV-I-infected individuals (asymptomatic individuals, patients with adult T-cell leukemia, or patients with HTLV-associated myelopathy [TSP/HAM]) through the courtesy of W. Blattner. These sera were tested by indirect immunofluorescence for Tax, Rex, Rof, or Tof reactivity. Twenty-one of 28 serum samples were

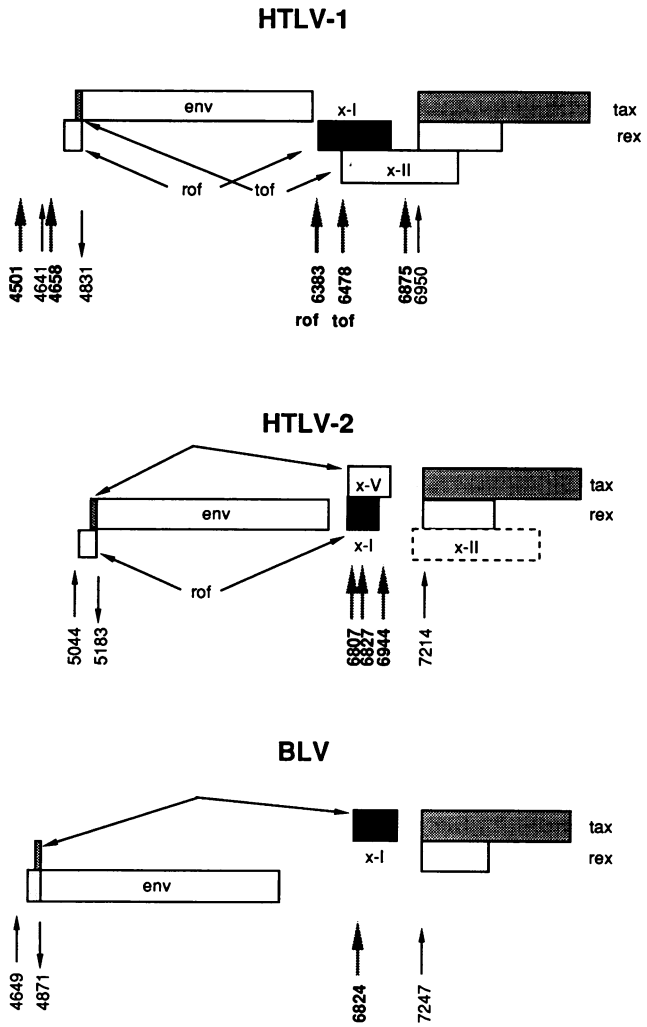

FIG. 5. Comparison of the X regions of different viruses. The organization of ORFs and identified splice sites of HTLV-I, HTLV-II, and BLV is shown. Solid vertical arrows indicate splice sites. Stippled arrows indicate the splice sites identified in this work. Numbers indicate the exact positions of the splice sites. (Numbering of HTLV-I starts at the beginning of transcription, while numbering of HTLV-II and BLV starts at the beginning of the U3 region of the 5' LTR.) Regions of homology are indicated by the different types of shading.

positive for Tax. In contrast, none of the sera was positive for Rex, Rof, or Tof. This striking difference in reactivities suggests that Rex, Tof and Rof might be much less immunogenic than Tax. The reasons for this difference are not clear. Rex and Tof, as nucleolar proteins, may not be easily accessible to the immune system. Interestingly, antibodies against Rev of HIV-1, another nucleolar protein, have not been detected in the majority of HIV-1-infected individuals. Analysis of additional serum samples, especially those indicating reactivity to Rex, is necessary to verify the expression of Rof and Tof proteins in infected individuals.

DISCUSSION

All members of the HTLV family of retroviruses contain the X region located between the *env* terminator and the 3' LTR. The distal portion of the X region encodes the essential proteins Tax and Rex. The proximal portion of the X region contains ORFs of unknown importance and function. It is interesting that despite the compact arrangement of the genome, these retroviruses contain a long region (more than 500 nt) with no identified function. This is a unique case, with no precedence among the known retroviruses. A hypothesis explaining the conservation of the proximal part of the X region in these viruses is that the detected ORFs in this region express additional protein products.

Conventional cDNA cloning did not reveal the presence of RNAs expressing the proximal portion of the X region. The detailed study of HTLV RNAs by using the reverse transcriptase-PCR technique described in this work demonstrated the presence of additional 3' SS in this region. The produced doubly-spliced mRNAs can express proteins utilizing either the Rex or the Tax AUG, named Rof and Tof, respectively. On the basis of Northern blot experiments, we estimated that *rof* and *tof* mRNAs are approximately 20-fold less abundant than the *env* mRNA. It is possible that *tof* and *rof* mRNAs are expressed at higher levels at certain times after infection or in different cell types. It has been shown previously that the levels of expression of different mRNAs produced by HIV-1 vary greatly during the viral life cycle (1, 12, 14, 23).

With HTLV-I, one of these proteins, named Tof, is selectively targeted to the nucleolus. Therefore, HTLV-I produces two proteins that are targeted to the nucleolus, Rex and Tof. Tof contains regions rich in serine and threonine. Computer searches of the protein data bases revealed similarities between Tof and serine-rich proteins, including several transcriptional activators such as oct-1, oct-2, pit-1, engrailed, and POU-M1. The serine/threonine-rich regions have been shown to be necessary activation domains in these proteins (20, 45, 46). This analysis suggested that Tof may function as an activator of transcription. Our experiments indicated that Tof was not able to activate expression from the HTLV-I LTR promoter and that it did not affect the extent of transactivation by the Tax protein. It remains to be determined whether Tof can activate transcription from cellular promoters.

Rof contains the arginine-rich region important for the binding of Rex to RXRE. The rest of the protein is predicted to be highly hydrophobic and shows homology to hydrophobic domains of several known membrane proteins. It is not known whether Rof binds to RXRE or whether it is localized in the nucleolus. Our experiments indicate that Rof does not interfere with Rex function.

It has been shown that the regulatory and accessory proteins of HIV-1 and the other lentiviruses are produced by additional alternatively spliced mRNAs (3, 4, 7, 28, 36–38, 47). Complex splicing results in the ability to express many proteins from one promoter. Our results indicate that the HTLV family also makes extensive use of this strategy to express additional proteins, which may play a role in the viral life cycle or interact with and modify cellular functions. The role of these new protein products in the viral life cycle and in the pathogenesis of disease needs to be examined in more detail. The availability of a full-length HTLV-I molecular clone described here, which is able to express all viral proteins and generates viral particles, will be helpful for future experiments.

ACKNOWLEDGMENTS

We thank S. Schwartz for assistance with experiments and for discussions, D. D'Agostino for discussions, M. Campbell and J. Harrison for excellent technical assistance, T. Copeland for assis-

tance with the antibody development, K. Nagashima and M. Gonda for electron microscopy, M. Powers for oligonucleotide synthesis, and A. Arthur for editing. We also thank W. Blattner and A. Manns for generously providing patient sera and I. Koralnik, M. Klotman, and G. Franchini for sharing unpublished data.

This research was sponsored in part by the National Cancer Institute, DHHS, under contract N01-CO-74101 with ABL.

REFERENCES

1. **Ahmad, N., R. K. Maitra, and S. Venkatesan.** 1989. Rev-induced modulation of nef protein underlies temporal regulation of human immunodeficiency virus replication. Proc. Natl. Acad. Sci. USA **86**:6111–6115.
2. **Ahmed, Y. F., S. M. Hanly, M. H. Malim, B. R. Cullen, and W. C. Greene.** 1990. Structure-function analysis of the HTLV-I rex and HIV-1 rev RNA response elements: insights into the mechanism of rex and rev function. Genes Dev. **4**:1014–1022.
3. **Arrigo, S. J., S. Weitsman, J. A. Zack, and I. S. Y. Chen.** 1990. Characterization and expression of novel singly spliced RNA species of human immunodeficiency virus type 1. J. Virol. **64**:4585–4588.
4. **Arya, S. K., and R. C. Gallo.** 1986. Three novel genes of human T-lymphotropic virus type III: immune reactivity of their products with sera from acquired immune deficiency syndrome patients. Proc. Natl. Acad. Sci. USA **83**:2209–2213.
5. **Benko, D. M., R. Robinson, L. Solomin, M. Mellini, B. K. Felber, and G. N. Pavlakis.** 1990. Binding of trans-dominant mutant rev protein of human immunodeficiency virus type 1 to the cis-acting rev responsive element does not affect the fate of viral mRNA. New Biol. **2**:1111–1122.
6. **Cann, A. J., J. D. Rosenblatt, W. Wachsman, N. P. Shah, and I. S. Y. Chen.** 1985. Identification of the gene responsible for human T-cell leukaemia virus transcriptional regulation. Nature (London) **318**:571–574.

6a. **Chen, I. S., J. McLaughlin, J. C. Gasson, S. C. Clark, and D. W. Golde.** 1983. Molecular characterization of genome of a novel T-cell leukemia virus. Nature (London) **305**:502–505.

7. **Colombini, S., S. K. Arya, M. S. Reitz, L. Jagodzinski, B. Beaver, and F. Wong-Staal.** 1989. Structure of simian immunodeficiency virus regulatory genes. Proc. Natl. Acad. Sci. USA **86**:4813–4817.
8. **Cullen, B. R.** 1991. Human immunodeficiency virus as a prototypic complex retrovirus. J. Virol. **65**:1053–1056.
9. **D'Agostino, D. M., B. K. Felber, J. E. Harrison, and G. N. Pavlakis.** The Rev protein of human immunodeficiency virus type 1 promotes polysomal association and translation of *gag/pol* and *vpu/env* mRNAs. Mol. Cell. Biol., in press.
10. **Derse, D.** 1987. Bovine leukemia virus transcription is controlled by a virus-encoded *trans*-acting factor and by *cis*-acting response elements. J. Virol. **61**:2462–2471.
11. **Derse, D.** 1988. *trans*-acting regulation of bovine leukemia virus mRNA processing. J. Virol. **62**:1115–1119.
12. **Feinberg, M. B., R. F. Jarrett, A. Aldovini, R. C. Gallo, and F. Wong-Staal.** 1986. HTLV-III expression and production involve complex regulation at the levels of splicing and translation of viral RNA. Cell **46**:807–817.
13. **Felber, B. K., D. Derse, A. Athanassopoulos, M. Campbell, and G. N. Pavlakis.** 1989. Cross-activation of the rex proteins of HTLV-1 and BLV and of the Rev protein of HIV-1 and nonreciprocal interactions with their RNA responsive elements. New Biol. **1**:318–330.
14. **Felber, B. K., C. M. Drysdale, and G. N. Pavlakis.** 1990. Feedback regulation of human immunodeficiency virus type 1 expression by the Rev protein. J. Virol. **64**:3734–3741.
15. **Felber, B. K., M. Hadzopoulou-Cladaras, C. Cladaras, T. Copeland, and G. N. Pavlakis.** 1989. rev protein of human immunodeficiency virus type 1 affects the stability and transport of the viral mRNA. Proc. Natl. Acad. Sci. USA **86**:1495–1499.
16. **Felber, B. K., H. Paskalis, C. Kleinman-Ewing, F. Wong-Staal, and G. N. Pavlakis.** 1985. The pX protein of HTLV-I is a transcriptional activator of its long terminal repeats. Science **229**:675–679.
17. **Gorman, C. M., L. F. Moffat, and B. H. Howard.** 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. **2**:1044–1051.
18. **Graham, F. J., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–460.
19. **Hadzopoulou-Cladaras, M., B. K. Felber, C. Cladaras, A. Athanassopoulos, A. Tse, and G. N. Pavlakis.** 1989. The *rev* (*trs/art*) protein of human immunodeficiency virus type 1 affects viral mRNA and protein expression via a *cis*-acting sequence in the *env* region. J. Virol. **63**:1265–1274.
20. **Ingraham, H. A., S. E. Flynn, J. W. Voss, V. R. Albert, M. S. Kapiloff, L. Wilson, and M. G. Rosenfeld.** 1990. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1 Pit-1 interactions. Cell **61**:1021–1033.
21. **Inoue, J., M. Seiki, and M. Yoshida.** 1986. The second pX product p27 chi-III of HTLV-1 is required for gag gene expression. FEBS Lett. **209**:187–190.
22. **Inoue, J., M. Yoshida, and M. Seiki.** 1987. Transcriptional (p40X) and post-transcriptional (p27X-III) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc. Natl. Acad. Sci. USA **84**:3653–3657.
23. **Kim, S., R. Byrn, J. Groopman, and D. Baltimore.** 1989. Temporal aspects of DNA and RNA synthesis during human immunodeficiency virus infection: evidence for differential gene expression. J. Virol. **63**:3708–3713.
24. **Kiyokawa, T., M. Seiki, S. Iwashita, K. Imagawa, F. Shimizu, and M. Yoshida.** 1985. p27XIII and p21XIII proteins encoded by the pX sequence of human T-cell leukemia virus type I. Proc. Natl. Acad. Sci. USA **82**:8359–8363.
25. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
26. **Longo, D. L., E. P. Gelmann, J. Cossman, R. A. Young, R. C. Gallo, S. J. O'Brien, and L. A. Matis.** 1984. Isolation of HTLV-transformed B-lymphocyte clone from a patient with HTLV-associated adult T-cell leukaemia. Nature (London) **310**: 506.
27. **Mann, D. L., J. Clark, M. Clarke, M. Reitz, M. Popovic, G. Franchini, C. D. Trainor, D. M. Strong, W. A. Blattner, and R. C. Gallo.** 1984. Identification of the human T cell lymphoma virus in B cell lines established from patients with adult T cell leukemia. J. Clin. Invest. **74**:56–62.
28. **Muesing, M. A., D. H. Smith, C. D. Cabradilla, C. V. Benton, L. A. Lasky, and D. J. Capon.** 1985. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature (London) **313**:450–458.
29. **Myers, G., B. Korber, J. A. Berzofsky, R. Smith, and G. N. Pavlakis (ed.).** 1991. Human retroviruses and AIDS. A compilation and analysis of nucleic acid and amino acid sequences. Los Alamos National Laboratory, Los Alamos, N.Mex.
30. **Myers, G., and G. N. Pavlakis.** 1991. Evolutionary potential of complex retroviruses, p. 1–37. *In* J. Levy (ed.), The retroviridae. Plenum Press, New York.
31. **Nagashima, K., M. Yoshida, and M. Seiki.** 1986. A single species of *pX* mRNA of human T-cell leukemia virus type I encodes *trans*-activator $p40^x$ and two other phosphoproteins. J. Virol. **60**:394–399.
32. **Paskalis, H., B. K. Felber, and G. N. Pavlakis.** 1986. Cis-acting sequences responsible for the transcriptional activation of human T-cell leukemia virus type I constitute a conditional enhancer. Proc. Natl. Acad. Sci. USA **83**:6558–6562.
33. **Pavlakis, G. N., S. Schwartz, D. M. Benko, and B. K. Felber.** 1991. Structure, splicing, and regulation of expression of HIV-1: a model for the general organization of lentiviruses and other complex retroviruses, p. 41–63. *In* R. Kennedy, F. Wong-Staal, and W. C. Koff (ed.), Annual review of AIDS research. Marcel Dekker Inc., New York.
34. **Rosen, C. A., J. G. Sodroski, R. Kettman, and W. A. Haseltine.** 1986. Activation of enhancer sequences in type II human T-cell leukemia virus and bovine leukemia virus long terminal repeat by virus-associated *trans*-acting regulatory factors. J. Virol.

57:738–744.
35. **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.
36. **Schwartz, S., B. K. Felber, D. M. Benko, E. M. Fenyö, and G. N. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J. Virol. **64**:2519–2529.
37. **Schwartz, S., B. K. Felber, E. M. Fenyö, and G. N. Pavlakis.** 1990. Env and Vpu proteins of human immunodeficiency virus type 1 are produced from multiple bicistronic mRNAs. J. Virol. **64**:5448–5456.
38. **Schwartz, S., B. K. Felber, and G. N. Pavlakis.** 1991. Expression of human immunodeficiency virus type-1 vif and vpr mRNAs is Rev-dependent and regulated by splicing. Virology **183**:677–686.
39. **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.
40. **Seiki, M., J. Inoue, M. Hidaka, and M. Yoshida.** 1988. Two cis-acting elements responsible for posttranscriptional trans-regulation of gene expression of human T-cell leukemia virus type I. Proc. Natl. Acad. Sci. USA **85**:124–128.
41. **Seiki, M., J. Inoue, T. Takeda, and M. Yoshida.** 1986. Direct evidence that the p40x of human T-cell leukemia virus type-1 is a trans-acting transcriptional activator. EMBO J. **5**:561–565.
42. **Shimotohno, K., W. Wachsman, Y. Takahashi, D. W. Golde, M. Miwa, T. Sugimura, and I. S. Y. Chen.** 1984. Nucleotide sequence of the 3′ region of an infectious human T-cell leukemia virus type II genome. Proc. Natl. Acad. Sci. USA **81**:6657.
43. **Sodroski, J., M. Trus, D. Perkins, R. Patarca, F. Wong-Staal, E. Gelman, R. Gallo, and W. Haseltine.** 1984. Repetitive structure in the long-terminal repeat element of a type II human T cell leukemia virus. Proc. Natl. Acad. Sci. USA **81**:4617–4621.
44. **Solomin, L., B. K. Felber, and G. N. Pavlakis.** 1990. Different sites of interaction for Rev, Tev, and Rex proteins within the Rev-responsive element of human immunodeficiency virus type 1. J. Virol. **64**:6010–6017.
45. **Tanaka, M., and W. Herr.** 1990. Differential transcriptional activation by oct-1 and oct-2: interdependent activation domains induce oct-2 phosphorylation. Cell **60**:375–386.
46. **Theill, L. E., J. W. Castrillo, D. Wu, and M. Karin.** 1989. Dissection of functional domains of the pituitary-specific transcription factor GHF-1. Nature (London) **342**:945–948.
47. **Viglianti, G. A., P. L. Sharma, and J. I. Mullins.** 1990. Simian immunodeficiency virus displays complex patterns of RNA splicing. J. Virol. **64**:4207–4216.