JOURNAL OF VIROLOGY, Feb. 1991, p. 727–735
0022-538X/91/020727-09$02.00/0

Vol. 65, No. 2

Analysis of Splicing Patterns of Human Spumaretrovirus by Polymerase Chain Reaction Reveals Complex RNA Structures

WALTER MURANYI AND ROLF M. FLÜGEL*

Project Group Human Retroviruses, Institut für Virusforschung, German Cancer Research Center, 6900 Heidelberg, Federal Republic of Germany

Received 29 August 1990/Accepted 14 November 1990

Mapping of transcripts of the human foamy virus genome was carried out in permissive human embryonic fibroblast cells by Northern blot hybridization and S1 nuclease analysis. Since several splice sites that are localized within a relatively narrow genomic region were detected, the polymerase chain reaction (PCR) was employed, and cloning and sequencing of the splice site junctions of the corresponding viral cDNAs were subsequently performed. All spumavirus transcripts have a common but relatively short leader RNA. Genomic, singly spliced *env* mRNAs and several singly and multiply spliced subgenomic transcripts were identified. The multiply spliced viral mRNAs consist of various exons located in the central or 3′ part of the viral genome. At least four novel gene products, termed Bet, Bes, Beo, and Bel3, are predicted to exist. The poly(A) addition site that defines the boundary of the R and U5 region in the 3′ long terminal repeat was determined. The pattern of spumavirus splicing is more complex than that of oncoviruses and more similar to that of lentiviruses. One of the characteristic features of spumavirus transcription is the existence of singly spliced *bel1* and *bel2* mRNAs that alternatively are multiply spliced, thereby generating a complexity comparable to, but different from, that of lentiviruses and from that of other known retroviruses. The complex spumavirus transcriptional pattern of human spumavirus and the coding potential of the 10 exons identified are discussed.

Foamy viruses, or spumaviruses, constitute a distinct group or subfamily of retroviruses that encode several novel genes, called *bel*, some of which may correspond to the regulatory genes *tat*, *rev*, and *nef* of the lentiviruses (10, 18, 21, 26). The *bel* genes are different from *gag*, *pol*, and *env* and located between the *env* and the 3′ long terminal repeat (LTR) of the human spumaretrovirus (HSRV) genome. We have previously reported on the molecular cloning and the primary structure of HSRV, which was originally isolated from a patient with nasopharyngeal carcinoma (1, 10, 15, 20). Study of the expression strategy of HSRV is important for understanding the life cycle and cytopathogenicity of the virus. Mapping of foamy virus transcripts of the foamy virus genome has not been conducted to date. To determine which genes are expressed after wild-type virus infection of human embryonic lung cells, transcription mapping was carried out by Northern (RNA) blot hybridization, S1 nuclease analysis, and polymerase chain reaction (PCR). Mapping and structural characterization of a surprisingly large number of HSRV transcripts gives insight into the expression and potential coding capacity of this subfamily of retroviruses.

MATERIALS AND METHODS

Preparation of RNA. Human embryonic lung fibroblast (HEL) cells were prepared and infections were performed as described previously (10). Total RNA was isolated from HSRV-infected HEL cells at different times points after infection, usually at the peak of the cytopathic effects. For most experiments, total RNA harvested after 3 and 7 days postinfection was prepared according to the guanidinium-thiocyanate procedure as described by Chirgwin et al. (4). Total RNA extracted from mock-infected HEL cells was prepared by the same procedure and run as a control.

* Corresponding author.

Northern blot hybridizations. A 10 to 20-μg sample of RNA was electrophoresed through a 1% agarose–2.2 M formaldehyde gel at a constant voltage of 80 V in a buffer consisting of 20 mM MOPS (morpholinepropanesulfonic acid), 5 mM sodium acetate, and 1 mM EDTA at pH 7.0. The gel was washed with distilled water, treated with 50 mM NaOH for 30 min, and subsequently neutralized in 100 mM Tris hydrochloride (pH 7.4). The RNA was transferred to a nylon membrane (GeneScreen Plus; NEN Corp.) overnight in 10× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate). The membrane was air dried, baked at 80°C, and prehybridized as suggested by the manufacturer. After 2 h the DNA probe, randomly labeled according to the method of Feinberg and Vogelstein (9), was added, incubated overnight at 42°C, and washed as recommended by the manufacturer. The nylon membrane was exposed to X-ray film at −70°C for 4 to 16 h.

S1 nuclease protection analysis. Appropriate DNA fragments of different regions of the HSRV genome were prepared from HSRV recombinant clones and labeled at either the 5′ or the 3′ terminus according to standard procedures (22). The viral DNA probes were heat denatured at 90°C and hybridized with 10 to 20 μg of RNA, depending on the percent G+C of the individual fragments (8). S1 nuclease digestions were performed with 50 U of S1 nuclease (Boehringer, Mannheim, Federal Republic of Germany) at 37°C for 1 h as described previously (22, 30). Since it is known that S1 nuclease cuts into A+T-rich regions at 37°C (3) and since the HSRV genome is relatively A+T rich (10), S1 nuclease digestions were also run at room temperature and at 4°C. The reaction products were ethanol precipitated and analyzed on a 5% polyacrylamide–8 M urea gel. The DNA fragments were used as probes and separately run as controls.

Synthetic oligodeoxynucleotide primers. Five sense primers and eight antisense primers were used. The following prim-

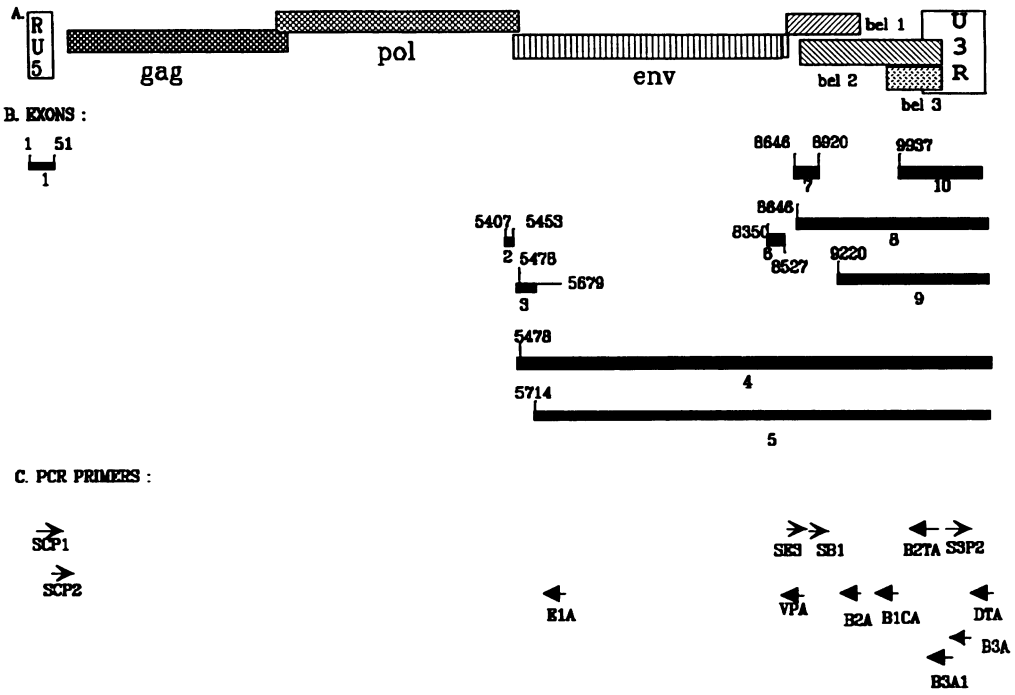

FIG. 1. (A) Organization of the HSRV genome. The viral ORFs are marked by differently hatched boxes. (B) The different exons are represented by solid bars and numbered continuously. Numbers indicate the exact positions of the exons in the HSRV RNA genome. (C) The positions of the amplification primers used are indicated and detailed in Materials and Methods. The splice acceptor of exon 10 was deduced from oligomer hybridization.

ers are in the sense strand: SCP1 and SCP2 (21-mers) are located in exon 1 of the HSRV genome at or downstream of the start site of transcription (15), SE3 (20-mer) is located in exon 4 of the *env* gene, SB1 (20-mer) is in exon 7 in the 5′ part of the *bet* gene, and S3P2 (18-mer) is located in exon 9. The following primers are all in the antisense strand: E1A (21-mer) is located in the 5′ part of exons 4 and 5; VPA (30-mer) is located in exon 6; B2A (20-mer) spans the splice acceptor of exon 9; B1CA (20-mer), B2TA (19-mer), and B3A1 (18-mer) are located within exons 8 and 9; B3A (21-mer) is in exon 10; and DTA (20-mer) spans the poly(A) addition site. The locations of the amplification primers are shown in Fig. 1. The sequences and the exact nucleotide positions of the primers in the RNA genome are as follows: SCP1, 5′-GCTCTTCACTACTCGCTGCGT-3′ (nucleotides [nt] 3 to 23); SCP2, 5′-CGAGAGTGTACGAGACTCTCC-3′ (nt 24 to 44); SE3, 5′-TGCAGCAACAAAGGACGTCT-3′ (nt 8474 to 8493); SB1, 5′-GGAGAGCTGACTATTGCT GA-3′ (nt 8751 to 8770); S3P2, 5′-GGTATGAGGTGTGTG GCT-3′ (nt 10601 to 10618); E1A, 5′-CCAGGCCAATACTC TTGAGCT-3′ (nt 5906 to 5926); VPA, 5′-AGCTAGCTGTT GTATCCAGGCAGGAGTATC-3′ (nt 8445 to 8474); B2A, 5′-GGACCTTCTGAGCAACTGGT-3′ (nt 9215 to 9234); B1CA, 5′-ATCGATGGATCGTCTCCTGG-3′ (nt 9312 to 9331); B2TA, 5′-GGTCCACCTGACATGTAAC-3′ (nt 9464 to 9482); B3A1, 5′-TGATGCCAAGCACCCTGC-3′ (nt 9885 to 9902); B3A, 5′-GAGGGAGTAGCGAATGAGGAC-3′ (nt 10188 to 10208); and DTA, 5′-$(T)_{10}$GGTTCTCGAA-3′ (nt 11012 to 11021). The primers SCP1 and SCP2 were used for amplification of all subgenomic multiply spliced mRNAs. The primer pair consisting of SCP1 and B1CA was used for amplification of cDNA of exons 1, 7, and 9; the primer combination of SCP1, SE3, and B1CA was used for exons 1, 6, 7, and 8; and the primer combination of SCP1, SE3, and VPA was used for exons 1, 3, and 6. The primer pair consisting of SCP1 and B2A was used for exons 1, 2, and 9; the primer pair consisting of SB1 and B3A was used for exons 6 and 10.

Reverse transcription of viral RNA and amplification of cDNAs by PCR. HSRV DNA was synthesized with Moloney murine leukemia virus reverse transcriptase for 1 h at 37°C or with avian myeloblastosis virus reverse transcriptase for 60 min at 41°C from 5 μg of total RNA from HSRV-infected cells under conditions specified by the manufacturers (GIBCO, Karlsruhe, Federal Republic of Germany, and Boehringer). Reverse transcriptions were started by adding an antisense oligodeoxynucleotide (150 pmol; 20-mer) derived from the 3′ region of the HSRV genome and random hexaoligodeoxynucleotides (0.5 μg per reaction; Pharmacia, Freiburg, Federal Republic of Germany) according to the method of Kinoshita et al. (14). Alternatively, reverse transcriptions were started with random hexanucleotide primers only, and the sense and antisense primers were added before the PCR to avoid amplification of false priming events. Concentrations of deoxynucleoside triphosphates were 2.5 mM in a reaction volume of 10 μl. The reaction products were denatured at 94°C for 5 min. Viral cDNAs were added to a solution that contained (in 90 μl) 0.25 mM concentrations of each of the four deoxynucleoside triphosphates, 150 pmol of a sense primer, and 10× *Taq* buffer. The PCR was started by adding 2 U of *Taq* DNA polymerase (Cetus Perkin-Elmer or Stratagene). Paraffin oil (50 μl [Merck]) was layered on top of the reaction mixture. The reaction mixture was incubated at 94, 55, and 72°C for 1, 2, and 3 min, respectively. The cycle was repeated 35 times in a DNA thermal cycler (Cetus Perkin-Elmer). Total RNA from unin-

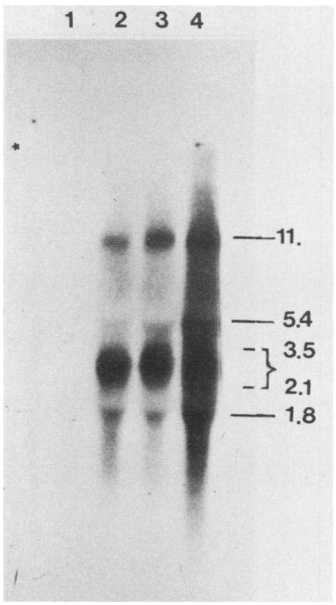

FIG. 2. Northern blot hybridization with the 430-bp-long *Hin*dIII-*Hin*dIII DNA subgenomic probe located in the 3′ part of the HSRV genome. The pockets, marked by an asterisk, were loaded with total RNA (20 μg) from mock-infected HEL cells (lane 1) or with total RNA (20 μg) from HSRV-infected cells 48 h (lanes 2 and 3) or 14 days after infection (lane 4). The numbers indicate the relative sizes of HSRV mRNAs in kilobases. The 11-kb mRNA species corresponds to genomic mRNA, the 5.4-kb species corresponds to singly spliced *env*, the 3.5- to 2.1-kb mRNAs correspond to singly and multiply spliced species of *bel1*, *bet*, and *bel2*, and the 1.8-kb transcript corresponds to singly spliced *bel2*. The time of exposure was 16 h.

fected HEL cells was run as a control under the same conditions but did not result in any HSRV-specific DNA bands. PCR products were analyzed on agarose gels by visualization with ethidium bromide or by hybridization to an oligodeoxynucleotide located in the exon. PCR-amplified DNA bands in the range of 150 to 500 bp were isolated by electrophoresis onto DEAE membranes (type NA45; Schleicher & Schuell), purified, and cloned into the polylinker site of a Bluescript vector according to standard methods (22). Larger and distinct DNA fragments were directly isolated from agarose gels and purified. Some viral DNAs were prechecked by restriction enzyme analysis. Nucleotide sequencing was performed by the method of Maxam and Gilbert or the dideoxy-mediated chain termination method of Sanger (22).

RESULTS

Northern blot analysis of viral RNAs. To initially map the viral transcripts, total RNA was extracted from HSRV-infected HEL cells by the guanidinium-thiocyanate method (4) and fractionated on denaturing gels. After electrophoresis, RNA was transferred to a nylon membrane and hybridized with a cDNA probe from the 3′ part of the HSRV genome. At least five species of viral mRNAs were detected, as shown in Fig. 2: a genomic RNA of 11.0 kb, subgenomic *env* mRNAs of approximately 5.4-kb, several subgenomic transcripts ranging in size from 2.1 to 3.4 kb, and an mRNA of 1.8 kb that was shown by S1 nuclease analysis and PCR to correspond to the singly spliced *bel2* transcript (see below and Table 1). It is noteworthy that the 2.1- to 3.4-kb viral transcripts were relatively abundant and independent of the time of harvest and the DNA probe used, leaving the possibility that other species of viral RNAs might be contained within this broad band. When DNA probes covering the central part of *env* or even extending into the carboxy-terminal region of *pol* were used, the small discrete mRNA species were still detectable (data not shown). This indicates that exons from these regions contribute to the small HSRV transcripts and that more precise methods for mapping that are capable of higher resolution had to be used.

S1 nuclease protection analysis of viral mRNAs. To characterize the viral transcripts in more detail, S1 nuclease experiments were performed with 5′- and 3′-labeled DNA probes derived from the central and 3′ part of the HSRV genome. When, for instance, and *Eco*RI-*Hpa*I DNA fragment of 1,080 bp (probe C [Fig. 3]) was labeled at the 5′ *Hpa*I site with polynucleotide kinase and [γ-^{32}P]ATP, hybridized to total RNA from HSRV-infected HEL cells and (separate-

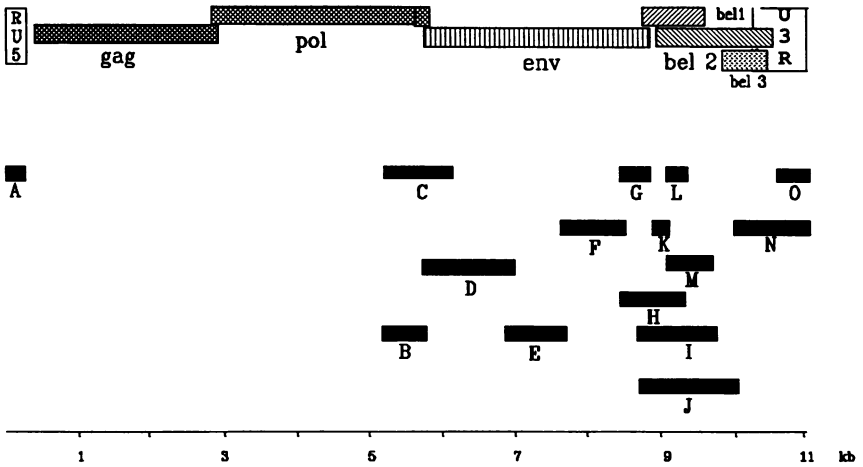

FIG. 3. Overview of subgenomic probes used for S1 nuclease protection analysis. The solid boxes show the locations and sizes of subgenomic probes. The probes are described in detail in Materials and Methods and in Table 1. Probes larger than 5 kb that correspond to recombinant plasmids described previously (10, 15) were omitted.

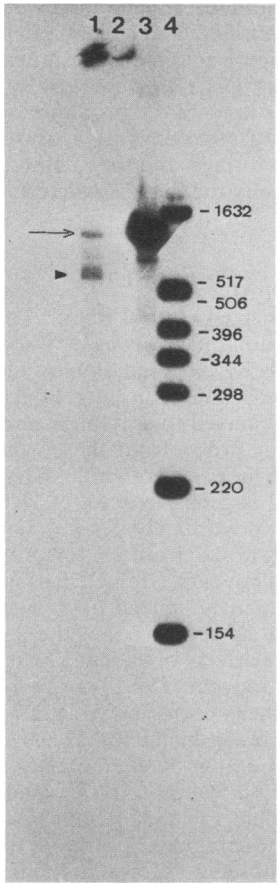

FIG. 4. S1 nuclease analysis of the splice acceptor of the HSRV *env* RNA. The 1,080-bp-long *Eco*RI-*Hpa*I probe C (Fig. 3) was labeled at the 5′ terminus and hybridized to 20 μg of total RNA from infected (lane 1) and uninfected (lane 2) HEL cells and digested with S1 nuclease as described in Materials and Methods. Lane 3 was loaded with an aliquot of probe DNA and lane 4 was loaded with *Hin*fI DNA fragments of pBR322 as markers (sizes given in base pairs on the right), and all of the lanes were analyzed on a 5% polyacrylamide–8 M urea gel. The thin arrow and the triangle mark protected fragments of 1,080 and 570 bp protected by genomic and *env* RNA, respectively. Note the faint band of 770 bp between the two intense bands that corresponds to the second *env* RNA transcript with a 5′ extension as discussed in Results.

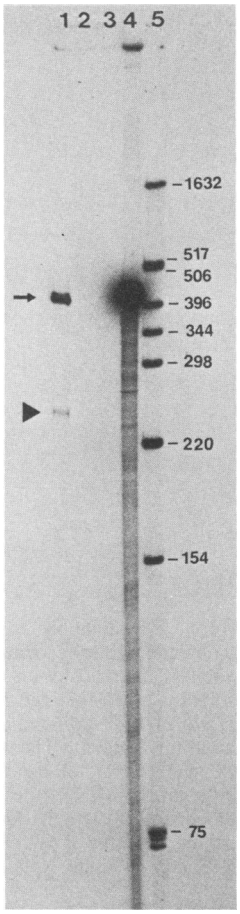

FIG. 5. S1 nuclease analysis of the splice acceptor of the *bell* mRNA. End-labeled probe H (Fig. 3) was hybridized to 20 or 2 μg of total RNA from infected HEL cells (lanes 1 and 2, respectively) and to 20 μg of total RNA from uninfected HEL cells (lane 3) and digested with S1 nuclease as described in Materials and Methods. Lane 4 contains an aliquot of probe H DNA, and lane 5 contains marker DNAs as described in the legend to Fig. 4. The arrow and triangle indicate 430- and 240-bp fragments protected by genomic and *bell* mRNAs, respectively.

ly) to RNA from mock-infected HEL cells, and subsequently digested with S1 nuclease, two bands of 1,080 and 570 bp protected against S1 nuclease were detected in RNA prepared from virus-infected but not from uninfected cells (Fig. 4). The first band corresponds to genomic RNA, whereas the second one derives from the *env*-specific transcript. This results in a splice acceptor site for the *env* mRNA around nt 5714 on the HSRV RNA genome map (15). In addition, a faint, 770-bp-long band migrating between the two intense bands was detected (Fig. 4, lane 1). This band was probably protected against S1 nuclease by an *env* transcript that has a 5′ extension of about 240 bp compared with the regular *env* transcript, which does not possess the 5′ extension. Its lower intensity could reflect the possibility that less of the second mRNA than of the *env* transcripts was present, since the same acceptor site is used by both exons 3 and 4, as shown by PCR (see below). The second *env* transcript encodes the same HSRV Env protein as the first *env* mRNA encodes.

Figure 5 shows that two bands of about 430 and 240 bp were protected against S1 nuclease digestion by an *Nhe*I-*Bam*HI DNA fragment (probe H [Fig. 3]); these bands correspond to genomic and to *bell* mRNA, respectively. This result locates the *bell* splice acceptor at approximately position 8646 on the HSRV RNA map. The splice acceptor is in excellent agreement with the canonical consensus sequences established for splice sites (24). Table 1 compiles the results obtained with a number of DNA probes used for S1 nuclease protection analysis. The sizes of the protected RNA exons allowed us to assign several splice sites to distinct genomic positions (Table 1) such as those of the *env*, *bet*, *bell*, and *bel2* transcripts. Those bands that were found to be protected against S1 nuclease digestion but could not be confirmed by PCR were additionally digested at a lower temperature. In these cases, it was found that the bands were not protected at 4°C, and close inspection of the neighboring sequences revealed that they were A+T rich. Consequently, these bands were not incorporated as splice sites into Table 1.

TABLE 1. End-labeled DNA probes for S1 nuclease protection analyses and assignment of splice sites

Restriction sites of DNA probe	Size (bp)	Labeled end	Size(s) of RNA(s) protected (bp)[a]	Genomic location(s) of splice sites[b]	Assignments of gene products[c]	Splice site confirmed by PCR
*Eco*RI, *Hpa*I[d]	1,080	5′	570	5714; SA	Env	Yes
*Eco*RI, *Hin*PI[d]	560	5′	360, 280	5407, 5478; SA	exon 3 (Bet)	Yes
*Nhe*I, *Bam*HI[d]	430	5′	240	8646; SA	Bel1	Yes
*Eco*RI,[d] *Bgl*II	1,080	3′	210	8920; SD	Bet, exon 7	Yes
*Eco*RI, *Hin*dIII[d]	1,220	5′	710	9220; SA	Bel 2	Yes
*Bam*HI, *Nco*I[d]	255	5′	145	9220; SA	Bel 2 (Bet)	Yes
*Bam*HI, *Hin*dIII[d]	814	5′	680	9220; SA	Bel 2 (Bet)	Yes
*Nhe*I,[d] *Cla*I	858	3′	58	8527; SD	Bet	Yes

[a] Protected RNA fragments derived from genomic RNA are not included.
[b] SA, Splice acceptor; SD, splice donor.
[c] Gene products in parentheses indicate subsequent ORFs as derived from PCR clones.
[d] Radiolabeled 3′ or 5′ terminus.

It was found that when DNA fragments prepared from recombinant clone B52 (10, 20) were used, a 137-bp-long DNA sequence located proximal to the 5′ end of the 3′ LTR did not protect any spumavirus RNAs against S1 nuclease digestion. This result was not obtained with any of the other foamy viral recombinant plasmids that did protect viral RNA. This indicates that only clone B52 contains a short insertion of nonviral DNA sequences, as previously assumed (15, 19). In view of the relatively high complexity of the pattern of HSRV transcription, it was necessary to employ the PCR method, which allows amplification of viral cDNAs present in minute amounts. Subsequent molecular cloning followed by nucleotide sequence analysis exactly determines the splice sites.

Synthesis of HSRV cDNA and amplification by PCR. To determine precisely the structure of the splice site junctions of subgenomic viral transcripts, cDNA was synthesized by using either murine leukemia virus or avian myeloblastosis virus reverse transcriptase from total RNA that had been prepared from HSRV-infected HEL cells. Oligonucleotides located in the 3′ or the central region of the virus genome served as primers during cDNA synthesis (Fig. 1). After heat denaturation, viral cDNAs were directly amplified in the same tube by adding *Taq* DNA polymerase and a sense primer located in the R region of the HSRV 5′ LTR immediately downstream of the transcription start site, as determined previously (15). Primers located in exon 1 and in the 3′ region were selected, since these exons are probably present in most viral transcripts. The PCR cycle was carried out as described in Materials and Methods and repeated 35 times. The PCR products were analyzed on agarose gels, stained with ethidium bromide, isolated, and purified according to standard techniques. Figure 6 illustrates the result of PCR products obtained in the presence and absence of random hexadeoxynucleotides. It is obvious that when random primers were present during reverse transcription and PCR, distinct and identifiable reaction products of viral origin were obtained with a lower overall background. In this case, all four DNA bands (marked in Fig. 6) were successfully cloned into the polylinker site of a Bluescript vector. Nucleotide sequencing of the viral inserts confirmed the origin and the presence of splice site junctions of two different *env* mRNa species and of *bet* and *bel2* transcripts. Table 2 presents the primer pairs that were used for the amplification of viral cDNAs and the resulting exon combinations that form the mature HSRV transcripts.

An essential result was the determination of the length of the leader RNA sequence that is common to all HSRV transcripts. The first splice donor site was found to be located at nt 51 downstream of the cap site with the RNA sequence UUG/GUU, consistent with donor site consensus sequences (24). This RNA leader represents exon 1, as shown schematically in Fig. 8, which contains an overview of the various exons identified. Although experiments with various oligomer primers located in the *gag* region were performed, only exon 1 was found.

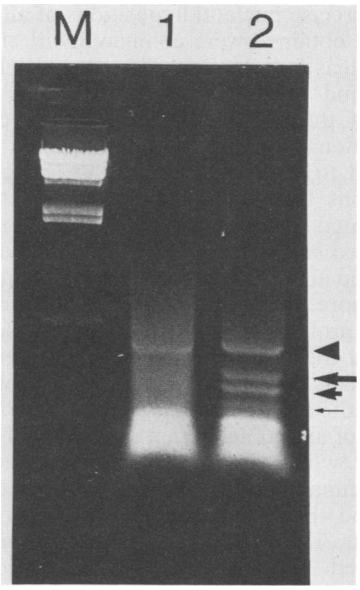

FIG. 6. PCR amplification of HSRV transcripts from viral RNA by reverse transcription and subsequent DNA amplification. Reverse transcription of 5 μg of total RNA from HSRV-infected HEL cells was done with primers E1A and B2A in the absence (lane 1) and in the presence (lane 2) of random hexamers as described in Materials and Methods (13). Viral cDNAs were subjected to 35 cycles of PCR amplification with *Taq* DNA polymerase in the presence of the sense primer SCP1. The PCR products were analyzed on a 1% agarose gel, visualized by staining with ethidium bromide, and isolated as described in Materials and Methods. The arrows mark the positions of viral cDNA fragments containing splice site junctions between exon 1 and exon 3 (the *env* mRNA species with a 5′ extension [solid triangle]), exon 1 and exons 6 and 7 (*bet* [thick long arrow]), exon 1 and exon 5 (*env* [thick short arrow]), and exon 1 and exon 9 (*bel2* [thin arrow]). Lane M contains lambda *Hin*dIII DNA fragments as markers.

TABLE 2. HSRV splice site junctions determined by PCR and nucleotide sequencing

Primer pair	Combination of exons identified	Assignment of mRNAs	Size of clone sequenced (bp)
SCP1 + E1A	1 and 5	*env*	300
SCP1 + E1A	1 and 4	*env*[a]	500
SCP1 + VPA	1, 3, and 6	*bet*	350
E2S + B1CA	6, 7, and 9	*bet*	450
SCP1 + B2A	1, 7, and 9	*bet*	350
SCP2 + B1CA	1, 7, and 9	*bet*	500
SCP1 + B2TA	1, 7, and 9	*bet*	600
SCP1 + B1CA	1 and 9	*bel2*	200
SCP1 + B3A1	1 and 8	*bel1*	1,300
SE3 + B1CA	6, 7, and 9	*bet*	300
SB1 + B1CA	7 and 9	*bet*	300
S3P2 + DTA	None	Poly(A) addition site	400
SCP1 + B2A	1, 2, and 9	*bel2*	150
SB1 + B3A	7 and 10[b]	*bel3*, *beo*	420

[a] This *env* transcript refers to the second mRNA shown in Fig. 4 and 6.
[b] This exon combination was identified by oligomer hybridization.

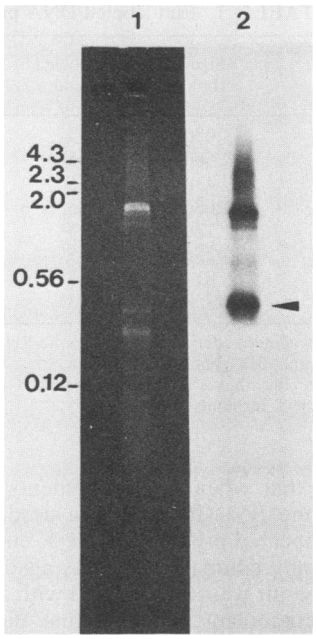

FIG. 7. PCR amplification of HSRV transcripts from viral RNA by reverse transcription and DNA amplification. Reverse transcription of 5 μg of total RNA from chronically infected HEL cells was done with random hexamer primers. Viral cDNAs were amplified as described in the legend to Fig. 6, except that primers SB1 and B3A were used. The PCR products were analyzed as described in the legend to Fig. 6 and probed with a labeled 20-mer, 5′-TCCTAGGA TTGGAGAAGGAC-3′ (nt 9963 to 9982). The arrow marks exons 7 and 10 of the *bel3* transcript of 420 bp. Numbers on the left indicate sizes of lambda *Hin*dIII DNA marker fragments in kilobases.

The determination of the primary structure of the viral cDNA inserts obtained by PCR revealed 10 different splice site junctions utilized in the generation of the various subgenomic transcripts, namely, those for the *env*, *bel1*, *bet*, *bel2*, and *bel3* gene products and for the noncoding exons 1, 2, 3, and 6, as shown in Table 2 and Fig. 1, which also includes the precise nucleotide positions of all exons identified. The data obtained were compared with the splice sites from S1 nuclease protection experiments. In almost all cases, this kind of analysis resulted in an unambiguous assignment of the different exons to their corresponding genomic locations. In some cases, restriction enzyme analysis was used to assign splice sites unambiguously to genomic locations that had been determined before by sequencing related molecular clones, e.g., exon 7 of the multiply spliced *bet* species (Table 2). The two *Bam*HI sites that are located in the *bel1* region were particularly useful. It was, furthermore, necessary to use a labeled oligodeoxynucleotide as a probe to identify the *bel3* transcript (Fig. 7), indicating that the *bel3* mRNA is not an abundant species. The splicing sites detected (Table 2) would join exon 7 and exon 10, i.e., the amino-terminal region of *bel1* provides the basic stretch of amino acids that is combined with the *bel3* coding region, which contains a leucine heptad repeat.

Figure 8 summarizes the salient features of the HSRV transcripts and the deduced protein-coding regions of the different exons. The *bel1* and *bel2* genes are either singly or multiply spliced. At least three of the exons identified have the potential to code for novel gene products (Fig. 9). These were termed *bes* and *bet*. The first is a truncated version that consists of the last 93 amino acid residues of the carboxy-terminal part of *bel1* (Fig. 9A). The corrected protein sequence (10a) of the putative *bel1* gene product is also shown in Fig. 9A. Bet encompasses 88 amino acid residues of the amino-terminal region of the *bel1* open reading frame (ORF) derived from exon 7 (Fig. 8), which is subsequently spliced into an acceptor site located in the central part of the *bel2* coding region, exon 8. This results in a putative protein of 480 amino acids for Bet (Fig. 9C). The deduced Bel2 protein sequence is shown in Fig. 9B.

Finally, it is possible that exon 7 is alternatively spliced into an acceptor site located in the *bel3* region (Table 2). Hybridization with oligodeoxynucleotide probes from the central part ruled out the possibility that exons 2 and 3 participate in the synthesis of the *bel3* mRNA. Two gene products, Bel3 and Beo, can be generated through this splicing; like Bet, they share the amino-terminal domain of the Bel1 protein (Fig. 8). The Bel3 protein sequence (Fig. 9E) is noteworthy, since it displays all of the hallmarks of a leucine zipper protein (29). The second protein, termed Beo (derived from Bel1 and Bel2) consists of the Bel1 amino-terminal domain and the carboxy-terminal part of Bel2 (Fig. 9D).

To determine the poly(A) addition site, an antisense 20-mer primer that consisted of an oligo(dT)$_{10}$ plus a decamer of the sequence located directly upstream of the presumed poly(A) site was used (primer DTA [Fig. 1]). A viral cDNA of 400 bp that upon nucleotide sequencing revealed the expected short stretch of poly(dA) that was covalently linked to the DNA sequence directly upstream of this site was obtained; thus, the foamy virus poly(A) addition site was unambiguously identified at position 11021 (Table 2). It is consistent with a recently published report on that of the closely related simian foamy virus type 1 (SFV-1) (17).

DISCUSSION

The results of mapping splice sites of HSRV transcripts reveal features that are common to and different from other known retroviruses. Starting with similarities, it was found that there is a leader RNA common to all HSRV transcripts. The size of HSRV exon 1 was determined to be 51 bases, which is surpringly short with respect to the overall size of

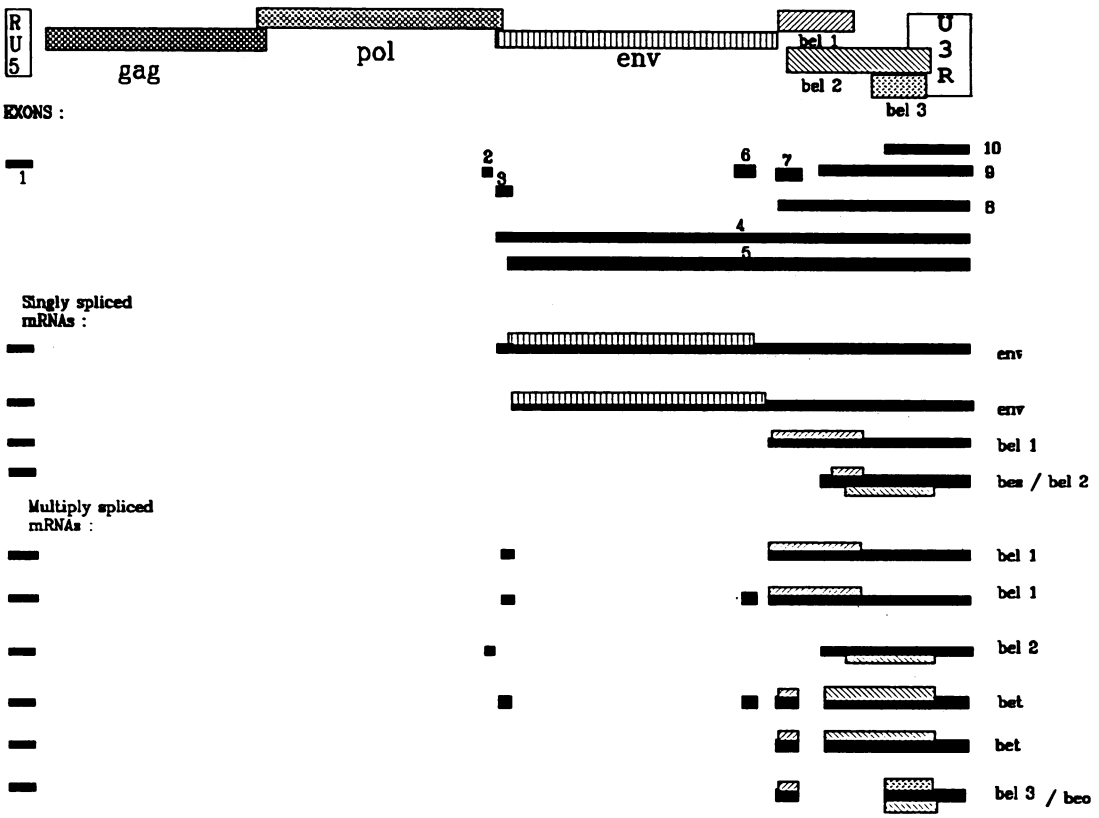

FIG. 8. Overview of the HSRV genomic organization and the corresponding exons and ORFs. The hatched boxes mark the different viral genes. The solid boxes indicate the different exons that are numbered from the 5′ end to the 3′ end. The singly and multiply spliced mRNAs contain ORFs that are marked by correspondingly hatched boxes to emphasize their coding capacities. Putative splicing of exon 7 to exon 10 was determined by oligomer hybridization.

the HSRV LTR. In fact, the HSRV exon 1 is the shortest leader RNA of all known retroviruses (31).

The spumavirus *env* mRNAs are singly spliced like those of all other retrovirus *env* messages. Three different splice acceptors were found upstream of the HSRV *env* transcript (Fig. 8). Two of the three upstream acceptor sites are utilized, so that two *env* mRNAs of slightly different size exist, similar to the *env* transcripts of visna virus (7). This was verified by S1 nuclease analysis and PCR. The first upstream splice acceptor, at position 5407 (Fig. 1), generates an mRNA that is spliced out at donor site 5453, and this exon 2 is then spliced into exon 9 for the expression of one of the *bel2* transcripts (Fig. 9). Exon 2 does not contain an AUG codon, and thus it does not seem to contribute to the protein-coding potential of the mature transcript.

A more complex transcriptional pattern was found with the acceptor at position 5478 which is common to exons 3 and 4 (Fig. 1). Splicing processes form an mRNA that may run through to the poly(A) addition site, thereby generating a second *env* transcript, exon 4; alternatively, the donor site at 5679 serves for the expression of exon 3 (Fig. 8). Exon 3 is subsequently spliced into exon 6, which is then spliced into exon 7, which expresses the first domain of the *bet* gene product (Fig. 9). This was demonstrated by sequencing a viral cDNA clone obtained by PCR. Again, as discussed above for exon 2, exons 3 and 6 do not encode an initiator; thus, a novel gene product does not appear to be expressed. This is in agreement with recent sequence data which show that the S1 regions of different HSRV recombinant clones do not possess an ORF. This result is also consistent with sequence data from the closely related SFV-1, which does not code for an S1 gene product (18). Another similarity exists between spumaviruses and lentiviruses in that the novel regulatory genes *tat* and *rev* of human immunodeficiency virus (HIV) and their putative spumavirus counterparts *bel1*, *bel2*, and *bet* of HSRV are multiply spliced (6, 23). In the case of the HSRV genes, some splicing events utilize exons positioned upstream from *env* or from within the *env* gene. But here the similarity stops, since all of the protein-coding potential of *bel1*, *bel2*, and *bet* stems from downstream of the HSRV *env* gene, in contrast to the protein-coding potential of the regulatory genes of the lentiviruses. Actually, in the *pol-env* overlap region, the splicing pattern is more similar to that of the human T-cell lymphotropic virus (HTLV) group of retroviruses.

The *bel1* splice acceptor is common to exons 7, 8, and 10. Exon 7 contains the amino-terminal part of the ORF for the first domain of the *bet* gene product. It is spliced out at donor site 8920 (Fig. 1) and joined to the acceptor site at position 9220. The resulting protein, Bet (for Bel1 and Bel two), has a remarkable sequence in that it joins features of the Bel1 sequence with that of the body of Bel2 (Fig. 9C).

Exon 8 contains the complete *bel1* coding region. Figure 9A shows the corrected sequence that has extensions at both the amino and carboxy termini (10, 10a). Although it has a segmental protein sequence homology to that of the transactivator protein Tat of HIV type 2 (16), *bel1* differs markedly from *tat* in that it can be singly spliced. There is

FIG. 9. ORFs deduced from sequencing PCR-amplified HSRV cDNAs. (A) Sequence of the *bel1* gene product. Putative nuclear localization signals are underlined. The initiator for the putative Bes protein sequence is in boldface. (B) Sequence of the *bel2* gene product. Putative nuclear localization signals are underlined. (C) Sequence of the *bet* gene product. The *bel1* domain is in italics. Putative nuclear or nucleolar localization signals are underlined. The arrowhead marks the splice site junction of the corresponding viral mRNAs. The Met initiator of Bel 2 that is present within Bet is boxed. (D) Sequence of the putative *beo* putative gene product. (E) Sequence of the *bel3* gene product. The leucine heptad repeats are marked by boldface underlining. Amino acid residues 122 to 210 in italics are homologous to the HIV type 2 Tat protein sequence, as reported previously (16).

evidence from transient expression assays that foamy viruses encode a transcriptional transactivator (13a, 18). Furthermore, the segmental but significant protein homology of the exon 1 domain of HIV type 2 *tat* to the *bel1* gene product (16) indicates that this is probably the HSRV transactivator. Consistent with this hypothesis is the finding that the Bel1 protein occurs in the nuclei of infected cells (unpublished data). Experiments with an infectious HSRV recombinant DNA clone into which small deletions in the *bel1* gene had been constructed completely abolished its infectivity, indicating that the *bel1* gene is essential for virus replication (14b). On the other hand, it is remarkable that HSRV also generates *bel1* and *bel2* transcripts that are singly spliced. This is a characteristic feature of spumaviruses that is shared by lentiviruses; it was just recently reported that certain lentivirus genes can be singly spliced (2, 6, 12, 28). It was reported that even within the lentivirus group, the splicing patterns have different degrees of complexity (6). It is, furthermore, noteworthy that the *bel2* transcripts are either singly or multiply spliced, apparently giving rise to the same gene products. In contrast, the *bel1* mRNAs in their multiply spliced form lead to the protein-coding sequence of *bet*. Figure 9 presents the five ORFs deduced from the mapping and sequencing data. A fascinating aspect of the Bel1 and Bet sequences is the presence of characteristic stretches of basic amino acids typical for nuclear localization (5, 13, 23). Since the first 88 amino acid residues of Bel1 are linked to three ORFs of the novel gene products, namely, Bet, Beo, and Bel3, the strongly basic residues from 47 to 54 are all contained within the amino-terminal regions of the HSRV gene products (Fig. 9A, C, D, and E). It is of interest that the entire Bel2 sequence is contained within the Bet sequence (Fig. 9B and C). Formally, it may be possible that Bel2 or Bet has a function similar to that of Rex of HTLV and Rev of the lentiviruses, which have signal sequences for nuclear or nucleolar localization (25, 27); however, the functions of Bel2 and Bet remain to be determined. Since analysis of retrovirus mutants has played a critical role in the understanding of the complex viral life cycle (11), site-directed mutagenesis will be an important tool in clarifying the functions of the different HSRV genes in transcription and replication.

Exon 9, which codes for the putative Bel2 protein, in addition can encode a truncated form of the Bel1 gene product, called Bes (for Bel1 short version). The first AUG initiator that occurs in any of the ORFs in the singly spliced *bel2* mRNA species would be the most likely candidate for *bes* expression, since other initiator codons are present in all other *bel1* and *bet* transcripts. It remains unknown from which *bel2* transcript *bes* is expressed, if it is expressed at all. Because of its bicistronic nature, exon 9 seems to be similar to the HTLV type I exon that encodes the Tax and Rex proteins (31). The use of centrally located exons and the presence of 3′ overlapping genes in HSRV is comparable to the situation in HTLV type I.

To detect *bel3*-specific mRNA, PCR experiments with different combinations of sense and antisense primers, the latter located in the *bel3* region, were carried out. It was found by oligomer hybridization that exon 7 could be spliced into exon 10. The two resulting novel gene products, Bel3 and Beo, have a common Bel1 domain, so that antisera against Bel1 would react with both Bel3 and Beo but reacted with proteins of molecular sizes corresponding to those of Bel1 and Bes (our unpublished data). The Bel3 protein sequence contains an interesting motif, a leucine zipper with a basic stretch located upstream of the leucine heptad repeat (29). We assume that Bel3 protein forms homo- or heterodimers or both and, in addition, binds to DNA. It is of interest in this context that SFV-1 does not possess an ORF for *bel3* (14a), indicating a significant difference between HSRV and SFV-1.

In contrast to lentiviruses, spumaviruses apparently do not encode viral proteins in the region between *pol* and *env* that are analogous to *vif*, *vpr*, *vpu*, and *vpx*, although two exons were identified in this region. It remains to be seen whether the various *bel* coding regions identified here take over similar functions.

What might be the advantage for HSRV to develop alternative mRNAs for the same protein sequence? Although experimental evidence is not available, it could be that the extra noncoding exons regulate the amounts of the various gene products. From a phylogenetic viewpoint, spumaviruses may represent a structure intermediate between the HTLV-bovine leukemia virus group and the lentiviruses.

The HSRV splice donor and acceptor sites are in general consistent with the general consensus sequence for splice sites in eucaryotic cells (24), with the exception of the splice donor 5453. However, similar anomalous splice sites were reported recently for HIV type 1 (23). The splice sites in the

HSRV genome identified here are all conserved in that part of the SFV-1 sequence that has been published (17).

ACKNOWLEDGMENTS

We thank G. Darai for the synthesis of PCR primers, Katrin Köhle for resequencing HSRV clones, Helmut Bannert for excellent technical assistance, Howard Temin for discussions, and Harald zur Hausen for support.

This research was supported by grant 323-4001-01KI8901/6 from the Bundesministerium für Forschung und Technologie.

ADDENDUM

The changes in the HSRV genome sequence that were identified by resequencing an infectious HSRV DNA clone were described previously (10a).

REFERENCES

1. **Achong, G., P. W. A. Mansell, M. A. Epstein, and P. Clifford.** 1971. An unusual virus in cultures from a human nasopharyngeal carcinoma. J. Natl. Cancer Inst. **42**:299–307.
2. **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.
3. **Berk, A. J.** 1989. Characterization of RNA molecules by S1 nuclease analysis. Methods Enzymol. **180**:334–342.
4. **Chirgwin, J. M., A. E. Przybyla, R. J. MacDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
5. **Cochrane, A. W., A. Perkins, and C. A. Rosen.** 1990. Identification of sequences important in the nucleolar localization of human immunodeficiency virus Rev: relevance of nucleolar localization to function. J. Virol. **64**:881–885.
6. **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.
7. **Davis, J. L., S. Molineaux, and J. E. Clements.** 1987. Visna virus exhibits a complex transcriptional pattern: one aspect of gene expression shared with the acquired immunodeficiency syndrome retrovirus. J. Virol. **61**:1325–1331.
8. **Davis, L. G., M. D. Dibner, and J. F. Battey.** 1986. S1 nuclease protection assay, p. 282–284. *In* Basic methods in molecular biology, 1st ed. Elsevier Science Publishing, Inc., New York.
9. **Feinberg, A. P., and B. Vogelstein.** 1984. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. **137**:266–267.
10. **Flügel, R. M., A. Rethwilm, B. Maurer, and G. Darai.** 1987. Nucleotide sequence analysis of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes. EMBO J. **6**:2077–2084.

10a. **Flügel, R. M., A. Rethwilm, B. Maurer, and G. Darai.** 1990. Nucleotide sequence analysis of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes. EMBO J. **9**:3806. (Corrigenda.)

11. **Goff, S. P., and L. I. Lobel.** 1987. Mutants of murine leukemia viruses and retroviral replication. Biochim. Biophys. Acta **907**: 93–123.
12. **Guatelli, J. C., T. R. Gingeras, and D. D. Richman.** 1990. Alternative splice acceptor utilization during human immunodeficiency virus type 1 infection of cultured cells. J. Virol. **64**:4093–4098.
13. **Hauber, J., A. Perkins, E. P. Heimer, and B. Cullen.** 1987. Trans-activation of human immunodeficiency virus gene expression is mediated by nuclear events. Proc. Natl. Acad. Sci. USA **84**:6364–6368.

13a. **Keller, A., et al.** Submitted for publication.

14. **Kinoshita, T., M. Shimoyana, K. Tobinai, M. Ito, S.-I. Ito, S. Ikeda, K. Tajima, K. Shimotohno, and T. Sugimura.** 1989. Detection of mRNA for the tax_1/rex_1 gene of human T-cell leukemia virus type I in fresh peripheral blood mononuclear cells of adult T-cell leukemia patients and viral carriers by using the polymerase chain reaction. Proc. Natl. Acad. Sci. USA **86**:5620–5624.

14a. **Kupiec, J.-J.** Personal communication.

14b. **Löchelt, M.** Personal communication.

15. **Maurer, B., H. Bannert, G. Darai, and R. M. Flügel.** 1988. Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. J. Virol. **62**:1590–1597.
16. **Maurer, B., and R. M. Flügel.** 1987. The 3′ orf protein of human immunodeficiency virus 2 shows sequence homology with the *bel* 3 gene of the human spumaretrovirus. FEBS Lett. **222**:286–288.
17. **Mergia, A., K. E. S. Shaw, J. E. Lackner, and P. A. Luciw.** 1990. Relationship of the *env* genes and the endonuclease domain of the *pol* genes of simian foamy virus type 1 and human foamy virus. J. Virol. **64**:406–410.
18. **Mergia, A., K. E. S. Shaw, E. Pratt-Lowe, P. A. Barry, and P. A. Luciw.** 1990. Simian foamy virus type 1 is a retrovirus which encodes a transcriptional transactivator. J. Virol. **64**:3598–3604.
19. **Rethwilm, A., G. Baunach, K. O. Netzer, B. Borisch, and V. ter Meulen.** 1990. Infectious DNA of the human spumaretrovirus. Nucleic Acids Res. **18**:733–738.
20. **Rethwilm, A., G. Darai, A. Rösen, B. Maurer, and R. M. Flügel.** 1987. Molecular cloning of the genome of the human spumaretrovirus. Gene **59**:19–28.
21. **Ruben, S., A. Perkins, R. Purcell, K. Joung, R. Sia, R. Burghoff, W. A. Haseltine, and C. A. Rosen.** 1989. Structural and functional characterization of human immunodeficiency virus *tat* protein. J. Virol. **63**:1–8.
22. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 13.42–13.75. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
23. **Schwartz, S., B. K. Felber, D. M. Benko, E.-M. Fenyö, and G. Pavlakis.** 1990. Cloning and functional analysis of multiply spliced mRNA species of human immunodeficiency virus type 1. J. Virol. **64**:2519–2529.
24. **Shapiro, M. B., and P. Senapathy.** 1987. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. Nucleic Acids Res. **15**:7155–7174.
25. **Siomi, H., H. Shida, S. H. Nam, T. Nosaka, M. Maki, and M. Hatanaka.** 1988. Sequence requirements for nucleolar localization of human T cell leukemia virus type I pX protein, which regulates viral RNA processing. Cell **55**:197–209.
26. **Sodroski, J., W. C. Goh, C. Rosen, A. Dayton, E. Terwilliger, and W. Haseltine.** 1986. A second post-transcriptional transactivator for HTLV-III replication. Nature (London) **321**:412–417.
27. **Venkatesh, L. H., S. Mohammed, and G. Chinnadurai.** 1990. Functional domains of the HIV-1 *rev* gene required for trans regulation and subcellular localization. Virology **176**:39–47.
28. **Viglianti, G. A., P. L. Sharma, and J. I. Mullins.** 1990. Simian immunodeficiency virus displays complex patterns of splicing. J. Virol. **64**:4207–4216.
29. **Vinson, C. R., P. B. Sigler, and S. L. McKnight.** 1989. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. Science **246**:911–916.
30. **Weaver, R. F., and C. Weissmann.** 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: the 5′ termini of 15 S β-globin mRNA precursor and mature 10 S β-globin mRNa have identical map coordinates. Nucleic Acids Res. **7**:1175–1193.
31. **Weiss, R., N. Teich, H. Varmus, and J. Coffin (ed).** 1982. RNA tumor viruses, 2nd ed., p. 1059–1101. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.