JOURNAL OF VIROLOGY, May 1988, p. 1590–1597
0022-538X/88/051590-08$02.00/0

Vol. 62, No. 5

Analysis of the Primary Structure of the Long Terminal Repeat and the *gag* and *pol* Genes of the Human Spumaretrovirus

BERND MAURER,[1] HELMUT BANNERT,[1] GHOLAMREZA DARAI,[2] AND ROLF M. FLÜGEL[1]*

Institut für Virusforschung, Deutsches Krebsforschungszentrum, Postfach 101949,[1] and Institut für Medizinische Virologie, Universität Heidelberg,[2] 6900 Heidelberg, Federal Republic of Germany

Received 30 September 1987/Accepted 15 January 1988

The nucleotide sequence of the human spumaretrovirus (HSRV) genome was determined. The 5′ long terminal repeat region was analyzed by strong stop cDNA synthesis and S1 nuclease mapping. The length of the RU5 region was determined and found to be 346 nucleotides long. The 5′ long terminal repeat is 1,123 base pairs long and is bound by an 18-base-pair primer-binding site complementary to the 3′ end of mammalian lysine-1,2-specific tRNA. Open reading frames for *gag* and *pol* genes were identified. Surprisingly, the HSRV gag protein does not contain the cysteine motif of the nucleic acid-binding proteins found in and typical of all other retroviral gag proteins; instead the HSRV *gag* gene encodes a strongly basic protein reminiscent of those of hepatitis B virus and retrotransposons. The carboxy-terminal part of the HSRV *gag* gene products encodes a protease domain. The *pol* gene overlaps the *gag* gene and is postulated to be synthesized as a *gag/pol* precursor via translational frameshifting analogous to that of Rous sarcoma virus, with 7 nucleotides immediately upstream of the termination codons of *gag* conserved between the two viral genomes. The HSRV *pol* gene is 2,730 nucleotides long, and its deduced protein sequence is readily subdivided into three well-conserved domains, the reverse transcriptase, the RNase H, and the integrase. Although the degree of homology of the HSRV reverse transcriptase domain is highest to that of murine leukemia virus, the HSRV genomic organization is more similar to that of human and simian immunodeficiency viruses. The data justify classifying the spumaretroviruses as a third subfamily of *Retroviridae*.

Spumaretroviruses (foamy viruses) make up the third subfamily of exogenous retroviruses of animals and humans (for a review, see reference 49). Spumaviruses have been isolated from healthy animals and from animals with a variety of diseases. Bovine foamy virus was isolated from cattle with leukosis (20), feline foamy viruses were isolated from cats with malignant lymphomas (33), and simian foamy viruses were isolated from different monkeys with malignant lymphomas (30, 35). Human spumaretroviruses (HSRVs) have been isolated from various patients, including a patient with a nasopharynx carcinoma (1) (this is the virus used for this study), one case of chronic myeloid leukemia (52), one case of toxic encephalopathy (2), and five cases of subacute thyroiditis de Quervain (41, 50). Recently, a spumaretrovirus was isolated from a patient with non-A, non-B hepatitis after blood transfusion (28). No proven pathogenicity for any of the isolates has been reported (49). Seroepidemiological data indicate a natural prevalence of foamy virus infection among men in certain parts of Africa and the Pacific (19, 25). There is evidence for immunosuppression after onset of a persistent infection with foamy viruses in experimental animals (12). In cell culture, spumaretroviruses replicate rather poorly, but cause cytopathic effects with the formation of giant multinucleated cells and vacuoles in a wide range of cells (12).

In the absence of molecular biological data on the structure of the spumaretroviral genomes, we decided to molecularly clone spumaretroviral DNA and establish the primary structure of one of its members. The primary structure of the central and 3′ part of the HSRV genome that consists of the *env* and three novel genes, *bel1*, *bel2*, and *bel3*, has been reported recently (7). This report focuses attention on the structural elucidation of the 5′ long terminal repeat (LTR), the *gag* and *pol* genes, and on the phylogenetic relatedness of HSRV to the other subfamilies of *Retroviridae* as well as on the unique features that set spumaretroviruses apart from other known retroviruses.

* Corresponding author.

MATERIALS AND METHODS

Cells and virus. Cells of human embryonic lung fibroblasts (HEL) were prepared as described previously (5). Virus was kindly provided by P. Loh, and infection of HEL cells was done as described previously (7).

Construction of recombinant plasmids. The construction and establishment of recombinant plasmids harboring HSRV DNA inserts from HSRV-infected HEL cells has been described previously. Selection, amplification, and purification of the recombinant clones were done as described previously (7, 32).

DNA sequence analysis. DNA fragments were digested with restriction enzymes, purified by agarose gel electrophoresis, and labeled at their 3′ ends with the Klenow fragment of the *Escherichia coli* DNA polymerase I and an appropriate [α-^{32}P]deoxynucleoside triphosphate defined by the recognition sequence of the given endonuclease. Alternatively, T4 polynucleotide kinase and [γ-^{32}P]ATP were used for 5′ labeling. The labeled fragments were sequenced by the method of Maxam and Gilbert (23). Computer analysis of the sequence data was performed with the BSA program devised by S. Suhai at the German Cancer Research Center, Heidelberg.

Strong stop cDNA analysis (endogenous reaction). HSRV was purified by sucrose gradient centrifugation (7, 26). Concentrated virus (20 μl) was incubated for 1 h at 37°C in 50 mM Tris hydrochloride (pH 7.8)–2 mM $MnCl_2$–6 mM $MgCl_2$–10 mM dithiothreitol–0.02% Triton X-100–100 μM each dATP, dGTP, TTP–100 μCi of [α-^{32}P]dCTP (specific activity, 3,000 Ci/mmol)–100 μg of actinomycin D per ml.

Fifteen minutes after the start of the reaction, dCTP (nonradioactive) was added to a final concentration of 10 μM. The reaction was stopped by the addition of 20 mM EDTA, 0.5% sodium dodecyl sulfate, and 200 μg of pronase per ml (final concentrations). After phenol and $CHCl_3$ extractions, the ethanol-precipitated product was hydrolyzed with 0.3 M NaOH for 1 h at 50°C in the presence of 3 mM EDTA. After neutralization with 0.3 M HCl, the reaction products were precipitated with ethanol and run on a 1% agarose gel.

S1 nuclease protection analysis. The 5'-^{32}P-labeled *Hin*dIII-*Hin*PI fragment of 708 base pairs (bp) was purified by Sephadex G-75 chromatography. This DNA fragment was used as a probe and run as a control. After denaturation and strand separation, both strands of the *Hin*dIII-*Hin*PI fragment were separately hybridized to poly(A)$^+$ RNA from HSRV-infected HEL cells at either 42 or 52°C for 12 h. S1 nuclease digestion was done at 37°C for 40 min with 10 to 400 U of S1 nuclease by the method of Weaver and Weissmann (48). The reaction products were ethanol precipitated and analyzed on a 1% agarose gel.

Materials. Labeled compounds were purchased from Amersham Corp. (Arlington Heights, Ill.) and New England Nuclear Corp. (Boston, Mass.), and enzymes were from Boehringer Mannheim (Mannheim, Federal Republic of Germany) or New England BioLabs, Inc. (Beverly, Mass.).

RESULTS

Nucleotide sequence analysis of pSPHSRV-B-C11 DNA. The strategy used for determining the primary structure of pSPHSRV-B-C11 (C11) DNA and the restriction maps for some enzymes are given in Fig. 1. More than 95% of the DNA sequence was determined for both strands, and various multiple-cut restriction enzymes were used to verify the sequence in addition to those shown in Fig. 1, in particular, the enzymes *Dde*I, *Sau*3AI, *Taq*I, *Sty*I, *Ava*II, and *Nde*I. The sequence was determined for C11 DNA from positions 1 to 5340 to the *Hin*dIII site (Fig. 1). Thus, those recombinant clones that have been characterized previously (clones D2 and C55) do overlap clone C11 (32). The 3' *Hin*dIII site of Fig. 1 coincides with the 5' *Hin*dIII site of the 3' part of the HSRV DNA sequence reported previously (7).

The resulting nucleotide sequence of the viral DNA insert of the recombinant clone C11 is shown in Fig. 2 with the predicted protein sequences of the corresponding HSRV gene products. Open reading frames (ORFs) longer than 100 amino acid residues were not found on the opposite strand, consistent with other retroviral genomes. The DNA sequence presented in Fig. 2 includes the 5' LTR, the primer-binding site (PBS), the *gag* gene, and the main part of the *pol* gene. The main features of the sequence are discussed starting from the 5' end of this part of the HSRV DNA.

5' LTR sequence. The first 1,123 nucleotides of C11 DNA contain a number of key features of regulatory signals shown to be required for retroviral replication and transcription. To determine the length of the presumed RU5 of the HSRV sequence, we performed strong stop cDNA synthesis under conditions of an endogenous reverse transcriptase reaction using [α-^{32}P]dCTP. Detergent-disrupted HSRV was used as the source of viral RNA that served as the template, and the reaction products of reverse transcription were analyzed by alkaline hydrolysis and subsequent agarose gel electrophoresis. The experiment shown in Fig. 3A indicates that a DNA fragment of 350 ± 20 bp was resistant to NaOH hydrolysis (marked by arrow, lane 3). In addition, a broad but smaller DNA band of about 160 bp was also visible (lane 3) that was probably formed by either premature abrogation during cDNA synthesis or degradation of the product of strong stop cDNA synthesis. A control experiment (lane 2) showed that before the alkaline hydrolysis step, a fragment with lower mobility was formed in the endogenous reaction and that it was degraded by NaOH. This result demonstrates that RNA served as the primer in the initiation step of the HSRV

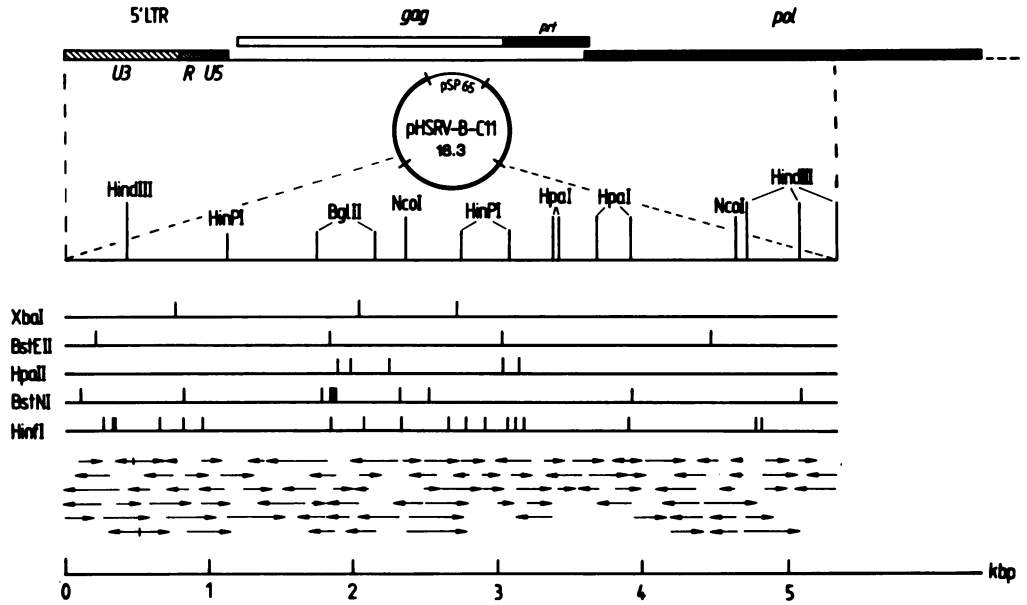

FIG. 1. Restriction maps of viral DNA insert of recombinant plasmid (pSPHSRV-B-C11) harboring HSRV-specific sequences and the strategy for determining the nucleotide sequence. The upper part shows the 5' part of the viral genome (5' LTR, *gag*, and *pol*). The arrows below the maps indicate the direction and extent of sequences determined for each fragment. The lower line represents the scale in kilobase pairs (kbp). The *pol* gene is in the first and the *gag* gene is in the second reading frame.

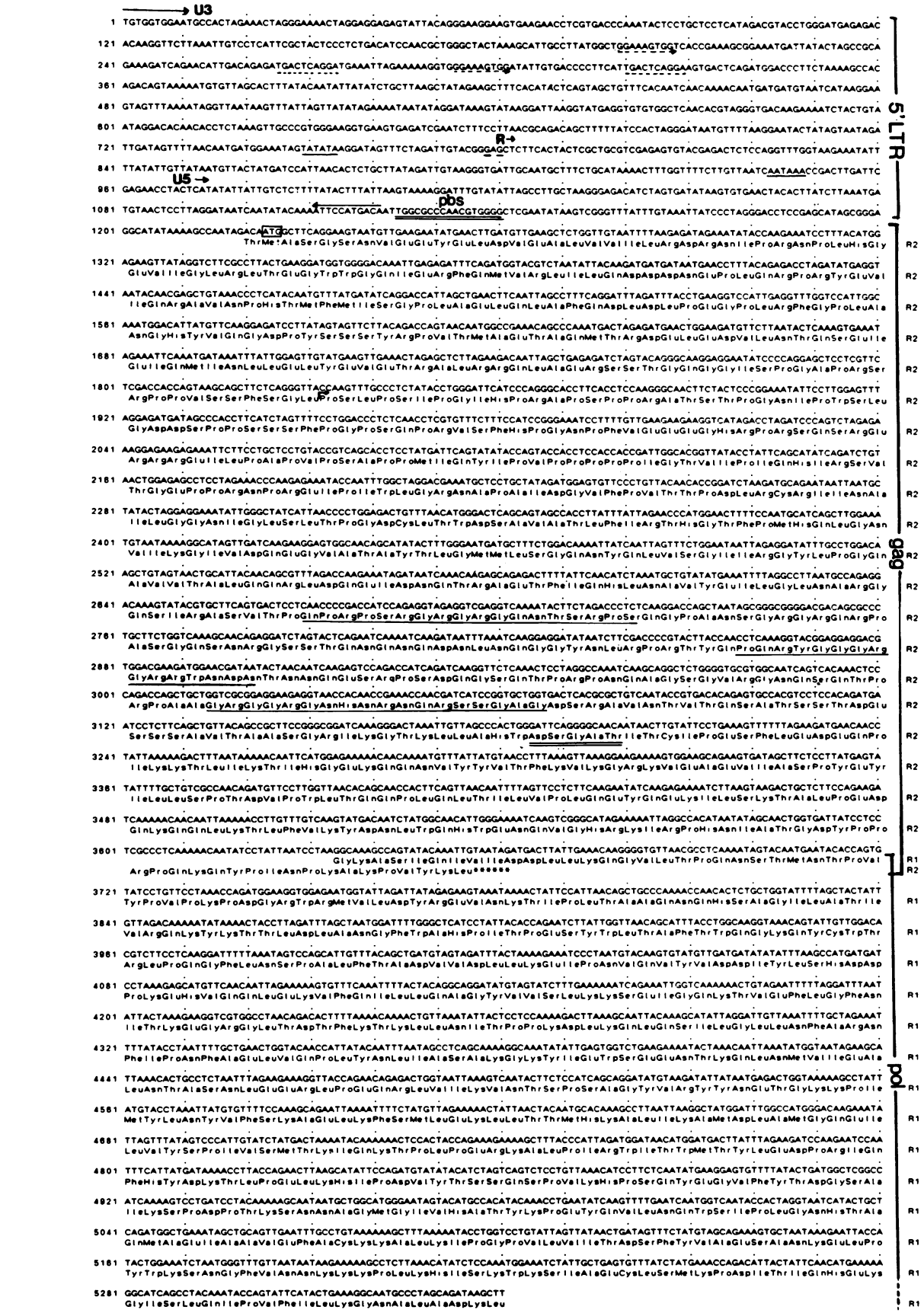

FIG. 2. DNA sequence of 5,340 bp of viral DNA insert of pSPHSRV-B-C11. The cap site, TATA box, and poly(A) addition signal are underlined. Direct repeats (AP-I-binding sites [17] and other putative enhancers) are marked by broken lines and arrows. The imperfect inverted repeat at the ends of the 5′ LTR is overlined by an arrow. The primer-binding site (pbs) is doubly underlined. The predicted amino acid sequences encoded by the *gag* and *pol* genes of HSRV are shown below the DNA sequence. One of the potential start codons for the *gag* protein is boxed. A small arrow (↱) indicates the NH_2 terminus of the major core protein. Some of the strongly basic stretches of the putative NABP are underlined. The conserved residues of the presumed HSRV protease are marked by double underlining. Stop codons are marked by asterisks.

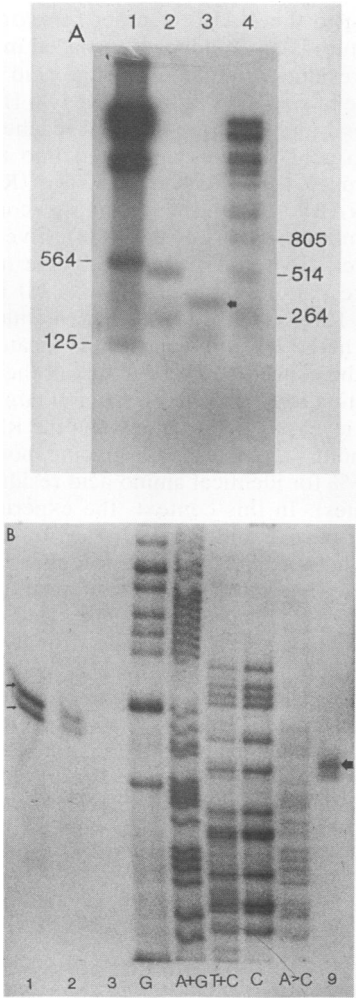

FIG. 3. (A) Strong stop cDNA synthesis. Endogenously synthesized strong stop cDNA was run on a 1% agarose gel before (lane 2) and after (lane 3) treatment with 0.3 M NaOH. *Hin*dIII DNA fragments (lane 1) and *Pst*I fragments (lane 4) from lambda DNA served as markers. The arrow is described in the text. (B) S1 nuclease protection analysis and strong stop cDNA synthesis. The experiments were performed as described in the legend for panel A. Lanes 1 to 3 were loaded with decreasing concentrations of the ^{32}P-labeled reaction products of the S1 nuclease analyses. The thin arrows mark the actual lengths of the protected DNA fragments from the *Hin*PI site within the PBS to the cap sites of the HSRV LTR. The sequence ladders were obtained by using the 708-bp *Hin*dIII-*Hin*PI DNA fragment that overlaps the RU5 region on both flanking regions. Lane 9 was loaded with the ^{32}P-labeled product of the strong stop cDNA synthesis. The thick arrow marks the reaction products of strong stop cDNA synthesis that are 346 and 344 ± 1 bp in size.

endogenous reverse transcriptase reaction, as shown before for other retroviruses (6, 9, 10, 42, 47).

To determine precisely the location of the cap site, ^{32}P-labeled strong stop cDNA was synthesized as described above, extracted, and analyzed on a 5% polyacrylamide gel. The lengths of the reaction products of strong stop cDNA synthesis were determined by electrophoresing them next to DNA sequence ladders obtained from the *Hin*dIII-*Hin*PI DNA fragment (central lanes in Fig. 3B) on the same gel. Two DNA bands were detected (lane 9 in Fig. 3B), one of which was predominant. According to this result, lengths of 346 and 344 ± 1 bp were obtained for RU5. These values are consistent with those obtained from the S1 nuclease protection analysis (see below). Furthermore, this result enables us to assign the cap sites to nucleotide positions 778 and 780 in the sequence shown in Fig. 2. To confirm this result, we performed S1 nuclease protection experiments using a 708-bp *Hin*dIII-*Hin*PI DNA fragment of clone C11 DNA that was hybridized to HSRV RNA. After S1 nuclease treatment, only one of the two strands hybridized to poly(A)$^+$ HSRV RNA. A nuclease-resistant fragment of about 350 bp was obtained (data not shown). This value reflects the size of the DNA-RNA hybrid that is protected against S1 nuclease digestion. This size is consistent with the size found by strong stop cDNA synthesis as described above.

It follows that the 5′ LTR of the HSRV has to be subdivided into a U3 region of 777 nucleotides and a R plus U5 region of 346 nucleotides. The cap site is preceded by a multiple but perfect TATATA box at −31 and followed by a consensus poly(A) signal sequence at positions 943 to 948. From analogy to the LTRs of other retroviruses, this results in a length of R = 196 and of U5 = 150 nucleotides (Fig. 2). The U3 region contains some direct and indirect repeats, some of which resemble enhancer elements found in other retroviruses (4, 34). The 3′ boundary of the 5′ LTR is defined by a stretch of 18 nucleotides that are complementary to the corresponding nucleotides at the 3′ end of the lysine-1,2-specific tRNA of mammals (29). The highest degree of similarity of the PBS, the primer for minus-strand HSRV DNA synthesis, is to that of visna virus (39).

Most retroviral LTRs are bounded by a perfect terminal inverted repeat of 11 to 12 bp; however, the HSRV LTR is bounded by a perfect inverted repeat of only 3 bp, or an imperfect repeat of 11 bp with nine identities similar to that of simian immunodeficiency virus (11) (Fig. 2). There are several other repeats, either direct or indirect ones, that have been marked in Fig. 2 the precise function of which is not known as yet, but they could play a role during integration (46). There are two nucleotides, AT, between the 3′ end of the 5′ LTR and the PBS (Fig. 2). The 5′ LTR and its 3′ adjacent region exhibit features analogous to those of other retroviruses, particularly to the lentiviral LTRs. Compared with the HSRV 3′ LTR (7), there are a number of point mutations, deletions, and insertions in the upstream LTR. The greatest difference was an insertion of 158 nucleotides in the downstream LTR between nucleotides 510 and 511 (Fig. 2). At exactly the same position, the 5′ LTR has instead 21 nucleotides that differ completely from the downstream LTR. Since both LTR sequences were derived from recombinant clones established from viral DNA, the step at which the insertion (or deletion) occurred remains obscure. However, the presence of a tetranucleotide sequence, TTTA, that occurs at both ends of the 158-nucleotide insertion of the 3′ LTR (7) argues for an insertion into the downstream LTR.

***gag* gene.** Downstream of the PBS of HSRV, precisely 99 nucleotides from the 3′ boundary of the 5′ LTR, a long uninterrupted reading frame of 2,436 nucleotides runs to two terminators at positions 3656 to 3661 (Fig. 2). The first start codon (ATG) of this ORF is located two triplets in, and its flanking sequences conform to those of a typical initiator (15). Consequently, the HSRV *gag* precursor would have a molecular size of 90 kilodaltons, slightly larger than those of most of other retroviruses. It is assumed that the *gag* precursor and the *gag/pol* precursor are processed by the retrovirus-encoded protease. The HSRV protease is presumably located in the COOH-terminal region of the *gag* gene, since there is some similarity between other retroviral pro-

QPRPSRGRGRGQNTSRPS	HSRV(1)
PQRYGGGRGRRWNDNTNN	HSRV(2)
GRGGRGNHNRNQRSSGAG	HSRV(3)
RDGNQGQRPGKGLSSGPW	RSV(3)
GLCYTCGSPGHYQAQCPK	RSV(1)
ERCQLCNGMGHNAKQCRK	RSV(2)
DQCAYCKEKGHWAKDCPK	MLV(1)
GCCFKCGKKGHFAKNCHE	MPMV(1)
GLCPRCKRGKHWANECKS	MPMV(2)
FKCWNCGKEGHSARQCRA	HIV-2(1)
QGCWKCGKPGHIMTNCPR	HIV-2(2)
QPCFRCGKVGHWSRDCTQ	HTLV-2(1)
GPCPLCQDPSHWKRDCPQ	HTLV-2(2)

FIG. 4. Amino acid sequences within the NABP of retroviral *gag* genes. Sequences duplicated within any one NABP are denoted (1) and (2). The region denoted as RSV (3) is located immediately downstream of that of RSV (2). Identities are marked by asterisks. The one-letter code for abbreviating amino acids is used.

tease sequences (45) and this limited region of the HSRV *gag* gene (Fig. 2). Furthermore, some of the conserved sequences within the HSRV protease (Asp-Ser-Gly-Ala-Thr) can be aligned with the active site residues of cellular acid proteases (24, 45).

Certain unusual features are found in the protein sequence of the HSRV *gag* gene products. Unexpectedly, the retroviral nucleic acid-binding protein (NABP) with its characteristic and periodic cysteine motif (n, $n + 3$, $n + 13$, with glycine and histidine residues at positions $n + 7$ and $n + 8$, respectively) is not present in the HSRV *gag* gene sequence. Instead, a strongly basic protein sequence is identified at a genomic location (between the p30 and the HSRV protease) corresponding to the NABP of other retroviruses. Its predominant feature is an equidistantly spaced internal peptide repeat of 18 residues with similar amino acid residues (Fig. 4). The degree of similarity between the two HSRV peptides is 33.3% (lines 1 and 2 in Fig. 4), and it reaches 72.2% when similar amino acid residues are taken into account. Like most retroviruses, the Rous sarcoma virus (RSV) *gag* gene encodes an NABP (p12) that contains two complete sets of the cysteine motif (n, $n + 3$, $n + 13$) (36). Five residues after the second set, RSV NABP has a glycine motif with the same distance pattern (n, $n + 3$, $n + 13$) as the HSRV NABP. This RSV glycine motif (Fig. 4, line 3) was also identified in the HSRV *gag* gene and is located 39 residues upstream of the conserved DSGAT box of the protease, at a genomic position found in other retroviral *gag* genes. Figure 4 depicts the level of similarity between the RSV third set—the glycine motif—and the HSRV glycine motif (lines 3 and 4) that is 27.8% for identical amino acid residues (72.2% for similar residues). In this context, the experiments of Goff and Lobel (8) are relevant. They showed by linker insertion mutagenesis that the NABP of murine leukemia virus (MLV) tolerates these insertions, since the resulting virus was perfectly viable (8). The absence of the cysteine motif in the HSRV *gag* sequence emphasizes the strongly basic character of all retroviral NABP sequences, particularly in the absence of strong data that the cysteine residues are required for biological activity.

Another interesting feature in the primary structure of the HSRV *gag* and *pol* genes is in the overlap region between them. The *pol* gene of HSRV overlaps by 22 nucleotides the *gag* gene that ends with TAATAG (Fig. 2). Since seven nucleotides (ACAAATTGTAATAG) upstream

```
                10        20        30        40        50        60        70
                .         .         .         .         .         .         .
 1 GKASIQIV--IDDLLKQGVLTPQNSTMNTPVYPVPKPDGR-WRMVLDYREVNKTIPLTAAQNQHSAGILA HSRV
 1 LGIKPHI----QRLLDQGILVPCQSPWNTPLLPVKKPGTNDYRPVQDLREVNKRVEDIHPTVPNPYNLLS MOMLV
 1 LNDL------VSKALEAGHIEPYSGPGNNPVFPVKKPNGK-WRFIHDLRATNAITTTLTSPSPGPPDLTS HTLV-II
 1 LAAAQQL---VQEQLEAGHITESSSPWNTPIFVIKKKSGK-WRLLQDLRAVNATMVLMGALQPGLPSPVA MPMV
 1 LVALTQL---VEKELQLGHIEPSLSCWNTPVFVIRKASGS-YRLLHDLRAVNAKLVPFGAVQQGAPVLSA RSV
 1 IKALVEICTEMEKEGKISKIGPEN-PYNTPVFAIKKKDSTKWRKLVDFRELNKRTQDFWEVQLGIPHPAG HIV-1
   o o           o  o o o  oo *o*oo  o*  o  o*   *o* ' *           o o
   L A           L  G I P  SP NTPVF  KK  G  WR   DLR VN            G P      CONSENSUS

                .         .         .         .         .         .         .
68 TI-VRQKYKTTLDLANGFWAHPITPESYWLTAFTWQG-------KQNCWTRLPQGFLNSPALFT-ADVVD HSRV
67 GLPPSHQWYTVLDLKDAFFCLRLHPTSQPLFAFEWRD-FEMGISGQLTWTRLPQGFKNSPTLFDEALHRD MOMLV
64 LP-TALPHLQTIDLTDAFFQIPLPKQYQPYFAFTIPQPCNYGPGTRYAWTVLPQGFKNSPTLFEQQLAAV HTLV-II
67 IP--QGYLKIIIDLKDCFFSIPLHPSDQKRFAFSLPSTNFKEPMQRFQWKVLPQGMANSPTLCQKYVATA MPMV
67 LP--RGWPLMVLDLKDCFFSIPLAEQDREAFAFTLPSVNNQAPARRFQWKVLPQGMTCSPTICQLVVGQV RSV
70 LK--KKKSVTVLDVGDAYFSVPLDEDFRKYTAFTIPSINNETPGIRYQYNVLPQGWKGSPAIFQSSMTKI HIV-1
             o*o o oo  oo       o**o o      o  o  o o****  o**ooo
             LDL D FF  PL       FAFT P      P  R  W VLPQG  NSPTLF          CONSENSUS

                .         .         .         .         .         .         .
129 L-----LKEIPNVQVYVDDIYLSHDDPKEHVQQLEK-VFQILLQAGYVVSLKKSEIGQKTVEFLGFNI   HSRV
136 LADFRIQHPDLILLQYVDDLLLAATSELDCQQGTRA-LLQTLGNLGYRASAKKAQICQKQVKYLGYLL   MOMLV
133 LNPMRKMFPTSTIVQYMDDILLASPTNEELQQLSQL-TLQALTTHGLPISQEKTQQTPGQIRFLGQVI   HTLV-II
135 IHKVRHAWKQMYIIHYMDDILIAGKDGQQVLQCFDQ-LKQELTAAGLHIAPEKVQLQD-PYTYLGFEL   MPMV
135 LEPLRLKHPSLCMLHYMDDLLLAASSHDGLEAAGEE-VISTLERAGFTISPDKVQREP-GVQYLGYKL   RSV
138 LEPFRKQNPDIVIYQYMDDLYVGSDLEIGQHRTKIEELRQHLLRWGLTTPDKKHQKEP-PFLWMGYEL   HIV-1
    o   o   o      *o** ooo        o       o *   *   o  * o        o*  o
    L   R   P      YMDD LLA        Q       Q L   G   S  K Q        LG  L    CONSENSUS
```

FIG. 5. Region of homologous protein sequences of the reverse transcriptase domains of the *pol* gene products of HSRV, Mo-MLV, HTLV-II, MPMV, RSV, and HIV-1. Amino acid residues identical for all six retroviral reverse transcriptase domains are marked by asterisks and by open circles when five or four of six residues are identical. Gaps were introduced to maximize homology. The one-letter code for abbreviating amino acids is used. Boundaries were set with respect to the HSRV sequence from positions 3634 to 4203 (Fig. 2). The boundaries at the carboxy termini of all other retroviral reverse transcriptases were taken from reference 39.

TABLE 1. Sequence similarity of the reverse transcriptase domain of retroviruses[a]

Retrovirus	% Sequence identity with:							
	HSRV	Mo-MLV	HTLV-II	MPMV	RSV	HIV-1	HIV-2	VIV[b]
HSRV		34.3	30.6	28.4	28.7	29.5	26.9	29.8
Mo-MLV			34.6	31.0	33.0	28.1	29.8	26.3
HTLV-II				40.1	39.2	33.7	35.1	30.8
MPMV					48.5	33.3	36.2	35.5
RSV						38.6	36.7	33.5
HIV-1							63.2	51.0
HIV-2								50.0

[a] The percentage of sequence identity value shown for each pair of retroviruses was obtained from the pairwise amino acid sequence alignments.
[b] VIV, Visna virus.

of the double terminators of the HSRV *gag* sequence exactly match the corresponding nucleotides of the RSV *gag* sequence ACAAATTTATAG, and since Jacks and Varmus (13) showed that the RSV *pol* gene is expressed by translational frame shifting, we assume that the HSRV *gag/pol* products are synthesized in the same way. This assumption is supported by the fact that the nucleotides that surround the first ATG of the *pol* ORF are not those of a typical initiator codon (15).

***pol* gene.** The second long ORF is undoubtedly that of *pol*, since the well-conserved amino acid sequences of the reverse transcriptase, the RNase H, and the integrase (endonuclease) domain can be readily identified (Fig. 2 and 5) (7). This subdivision into three domains with identifiable enzymatic activities and an intermediate region, termed the tether, located between the reverse transcriptase domain and the RNase H region is clearly consistent with the suggested alignment of *pol* genes by Varmus (46) and Johnson et al. (14). The value deduced from the HSRV nucleotide sequence predicts for the *pol* protein 908 amino acid residues, corresponding to 102 kilodaltons which add up to 193 kilodaltons if the *gag/pol* precursor is synthesized by translational frameshifting. Subsequent processing by the HSRV protease would result in at least six different polypeptides, among them four *gag*-derived structural proteins and two *pol* gene products, the reverse transcriptase proper and the integrase.

The degrees of homology between the HSRV reverse transcriptase domain and those of other retroviruses ranged from 34.3% (Moloney MLV [Mo-MLV]), 31.0% (human T-cell leukemia virus type I [HTLV-I]), and 29.5% (human immunodeficiency virus type 1 [HIV-1]) to 28.4% Mason-Pfizer monkey virus [MPMV]) (Table 1). This relatively high homology to MLV is consistent with the fact that the cation requirement for the purified reverse transcriptase from foamy viruses is much more biased for Mn^{2+} than for Mg^{2+} in the endogenous reaction (18, 27). To compare the conserved part of the reverse transcriptase domain, we selected sequences from those retroviruses that belong to three subfamilies (so that at least one member of the oncoviruses, lentiviruses, and spumaviruses was included). In addition, reverse transcriptase sequences from members of other groups (RSV, MPMV) were also included in the comparison shown in Fig. 5. The comparison reveals that residues located in regions shown to be functionally essential (16) are conserved in all retroviruses including HSRV (31, 37–40). Surprisingly, however, it was necessary to introduce two major gaps into the HSRV sequence for this multiple sequence comparison to achieve optimum alignment. These HSRV specific gaps were 7 and 5 amino residues long and located very close to those residues proven to be required for enzymatic activity (16). The comparison shows that the HSRV reverse transcriptase domain is clearly related to those of other retroviruses but that it is distinct, indicating the particular place of HSRV as a member of the third subfamily of *Retroviridae*. This result was supported by an analysis in which a pairwise comparison between the *pol* region of HSRV and that of the other members of different retrovirus groups or subfamilies was performed (Table 1). The lentiviruses are clearly classified by their relatively high amino acid identity of over 50% as reported previously from different laboratories (3, 31, 39, 40). The HTLV-bovine leukemia virus group viruses also seem to constitute a subfamily of their own as judged by the high degree of homology of over 70% between individual members (3). The D-type viruses, exemplified by MPMV, can readily be grouped together as another subfamily with high homologies (~90%) among their members, such as squirrel monkey retrovirus and simian retrovirus types 1 and 2 (44). A fourth subfamily is formed by the oncoviruses with Mo-MLV as the prototype; AKR murine leukemia virus and reticuloendotheliosis virus among others obviously belong into this subfamily.

The overall genomic organization of HSRV is diagrammatically represented in Fig. 6. The total genome size as well as the three extra *bel* genes set HSRV apart from other established subfamilies and groups of *Retroviridae*. This report together with the first report (7) complete the nucleotide sequence analysis of the HSRV genome. Thus, the overall genetic organization of HSRV is 5′ LTR, *gag-prt*, *pol*, S1, *env*, *bel1*, *bel2*, *bel3*, 3′ LTR, and by and large it resembles

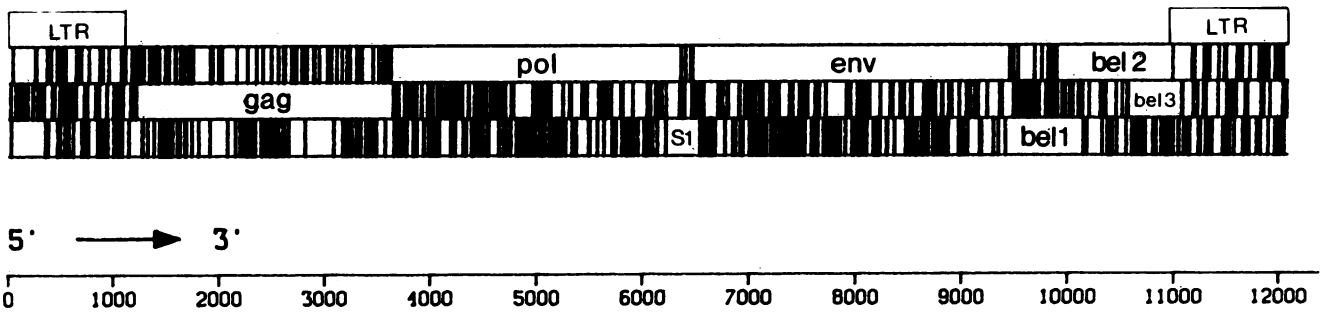

FIG. 6. Organization of the HSRV provirus. Vertical bars represent stop codons in the three reading frames. *bel1*, *bel2*, and *bel3* are the ORFs as described in reference 7. The boundaries of the 5′ LTR were taken from Fig. 2, and those of the 3′ LTR were from reference 7. Scale is in base pairs.

more those of the acquired immunodeficiency syndrome viruses than other retroviruses (51).

DISCUSSION

In the present studies, the nucleotide sequence of the 5′ part of the HSRV genome was determined and analyzed. The resulting primary structure of the HSRV DNA and the deduced amino acid sequences includes the 5′ LTR, the *gag* gene and the major part of the *pol* gene.

The 5′ LTR was characterized by determining the site of initiation of transcription by strong stop cDNA synthesis and S1 nuclease mapping. These analyses revealed that two start sites for viral transcription are used and placed the corresponding cap sites at nucleotides 778 and 780 of the 5′ LTR of HSRV. It is interesting that the U3 region of HSRV contains three repeats of the core consensus DNA sequence, TGACTCA, that is recognized by a recently identified human transactivator protein, a transcription factor termed AP-I (17), at −447, −459, and −514. Furthermore, there are additional direct repeat sequences in the U3 region that could serve as enhancer elements.

The 5′ LTR is followed by a PBS, downstream of which a leader sequence of 82 nucleotides precedes the first initiator codon of the *gag* protein sequence. The ORF for the HSRV *gag* protein comprises 811 amino acid residues. The protein sequence of the presumed *gag* precursor has several interesting features. Instead of the standard retroviral NABP with typical cysteine motifs, it contains regions of short, strongly basic domains that are predicted to bind to nucleic acids. A second distinct property of the HSRV *gag* protein is the presence of a proteaselike domain at the COOH terminus of the precursor. A similar genetic arrangement has been found in the RSV *gag* precursor.

The 3′ end of the HSRV *gag* gene overlaps the 5′ part of the *pol* gene by a short stretch of 22 nucleotides. Since part of this nucleotide region is identical to the corresponding site in the RSV genome for which a translational frameshift has been shown (13), we postulate a similar mechanism for the biosynthesis of the HSRV *gag/pol* precursor. As to the *pol* gene product itself, it is of great phylogenetic interest that the HSRV reverse transcriptase domain has a relatively high degree of homology to the corresponding sequences of other retroviruses, although it is, nonetheless, distinct.

Different criteria have been used to divide the family *Retroviridae* into three subfamilies, the oncoviruses, the lentiviruses, and the spumaretroviruses (21). This classification was mainly based on the length and sequence of strong stop cDNA (10), on the nature of the PBS, and on phylogenetic trees that were constructed by using either the reverse transcriptase or the integrase domain of retroviruses (3, 24, 39, 40).

With these criteria at hand and the data reported here on the structure of the HSRV genome, one expects that spumaretroviruses could be readily classified. However, problems that are inherent to the complex evolution of retroviruses have been emphasized recently and involved several different lineages connected by recombination (43).

At first sight, the HSRV reverse transcriptase and integrase domains have a higher degree of homology to those of the oncovirus MLV, apparently indicating that HSRV could belong to the oncovirus subfamily. However, the MLV genome is about 3 kilobases smaller, and it does not resemble at all that of HSRV in other regions, e.g., the LTR, *gag*, and *env* sequences.

On the other hand, the HSRV genome seems to encode potential new genes in its 3′ part. The *bel* genes have a limited protein similarity of 28 and 20% to the corresponding *tat* and 3′ *orf* gene products of HIV-2 (22). The presence of the extra *bel* genes argues for a relatedness to lentiviruses which is supported by the identity of 21 nucleotides of the HSRV PBS to the visna virus PBS (39). However, the reverse transcriptase domain of HSRV is clearly less related to that of the lentiviruses. For similar reasons, spumaretroviruses do not fit into the D-type and HTLV-bovine leukemia virus group of retroviruses. It is thus difficult to group foamy viruses into any of the previously established and characterized subfamilies or groups.

In conclusion, the criteria previously used for the classification of retroviruses lead to the contradictory result of placing HSRV into two different subfamilies, the lentiviruses and the oncoviruses. This paradox and the fact that the *gag* protein sequence of HSRV has unique features taken together with the overall genomic organization strongly indicate that HSRV is a member of a separate subfamily different from other subfamilies and groups of *Retroviridae*.

ACKNOWLEDGMENTS

We thank Howard Temin and Kunitada Shimotohno for critically reading the manuscript and M. Kasamasch for excellent preparation of the manuscript.

This project was supported by a research grant from the Deutsche Forschungsgemeinschaft (F1 134-1-1). The sequence data in this publication have been submitted to the EMBL/Genbank.

LITERATURE CITED

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. **Cameron, K. R., S. M. Birchall, and M. A. Moses.** 1978. Isolation of foamy virus from patient with dialysis encephalopathy. Lancet **ii**:796.
3. **Chiu, I.-M., A. Yaniv, J. E. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronick, and S. A. Aaronson.** 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. Nature (London) **317**:366–368.
4. **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.
5. **Flügel, R. M., B. Maurer, H. Bannert, A. Rethwilm, P. Schnitzler, and G. Darai.** 1987. Nucleotide sequence analysis of a cloned DNA fragment from human cells reveals homology to retrotransposons. Mol. Cell. Biol. **7**:231–236.
6. **Flügel, R. M., U. Rapp, and R. D. Wells.** 1973. RNA-DNA covalent bonds between the RNA primers and the DNA products formed by RNA tumor virus DNA polymerase. J. Virol. **12**: 1491–1502.
7. **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.
8. **Goff, S. P., and L. I. Lobel.** 1987. Mutants of murine leukemia viruses and retroviral replication. Biochim. Biophys. Acta **907**: 93–123.
9. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature (London) **326**:662–669.
10. **Haseltine, W. A., and D. G. Kleid.** 1978. A method for classification of 5′ termini of retroviruses. Nature (London) **273**:358–364.
11. **Hirsch, V., N. Riedel, and J. I. Mullins.** 1987. The genome organization of STLV-3 is similar to that of the AIDS virus except for a truncated transmembrane protein. Cell **49**:307–319.
12. **Hooks, J. J., and B. Detrick-Hooks.** 1981. Spumaviriniae: foamy virus group infections: comparative aspects and diagnosis, p. 599–618. *In* E. Kurstak and C. Kurstak (ed.), Comparative

diagnosis of viral diseases, vol. 4. Academic Press, Inc., New York.
13. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus *pol* gene by ribosomal frameshifting. Science **230:** 1237–1242.
14. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with nonviral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
15. **Kozak, M.** 1981. Possible role of flanking nucleotides in recognition of the initiator codon by eukaryotic ribosomes. Nucleic Acids Res. **9:**5233–5252.
16. **Larder, B. A., D. J. M. Purifoy, K. L. Powell, and G. Darby.** 1987. Site-specific mutagenesis of AIDS virus reverse transcriptase. Nature (London) **327:**716–717.
17. **Lee, W., P. Mitchell, and R. Tjian.** 1987. Purified transcription factor AP-1 interacts with TPA-inducible enhancer elements. Cell **49:**741–752.
18. **Liu, W.-T., T. Natori, K. S. S. Chang, and A. M. Wu.** 1977. Reverse transcriptase of foamy virus. Purification of the enzyme and immunological identification. Arch. Virol. **55:**187–200.
19. **Loh, P. C., F. Matsuura, and C. Mizumoto.** 1980. Seroepidemiology of human syncytial virus: antibody prevalance in the Pacific. Intervirology **13:**87–90.
20. **Malmquist, W. A., M. J. Van der Maaten, and A. D. Boothe.** 1969. Isolation, immunodiffusion, immunofluorescence, and electron microscopy of a syncytial virus of lymphosarcomatous and apparently normal cattle. Cancer Res. **29:**188–200.
21. **Matthews, R. E. F.** 1982. The classification and nomenclature of viruses. Intervirology **17:**234–238.
22. **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.
23. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74:**560–564.
24. **McClure, M. A., M. S. Johnson, and R. F. Doolittle.** 1987. Relocation of a protease-like gene segment between two retroviruses. Proc. Natl. Acad. Sci. USA **84:**2693–2697.
25. **Muller, H. K., G. Ball, M. A. Epstein, B. G. Achong, G. Lenoir, and A. Levin.** 1980. The prevalence of naturally occurring antibodies to human syncytial virus in East African populations. J. Gen. Virol. **47:**399–406.
26. **Obara, T., D. P. Bolognesi, and H. Bauer.** 1971. Ribosomal RNA in avian leukosis virus particles. Int. J. Cancer **7:**535–546.
27. **Parks, W. P., G. J. Todaro, E. M. Scolnick, and S. A. Aaronson.** 1971. RNA dependent DNA polymerase in primate syncytium-forming (foamy) viruses. Nature (London) **229:**258–260.
28. **Prince, A. M., B. A. A. Williams, T. Huima, L. Bardina, and B. Brotman.** 1984. Isolation of a virus from chimpanzee liver cell cultures inoculated with sera containing the agent of non-A, non-B hepatitis. Lancet **ii:**1071–1075.
29. **Raba, M., K. Limburg, M. Burghagen, J. R. Katze, M. Simsek, J. E. Heckman, U. L. Rajbhandary, and H. J. Gross.** 1979. Nucleotide sequence of three iso-accepting lysine tRNAs from rabbit liver and SV40-transformed mouse fibroblasts. Eur. J. Biochem. **97:**305–318.
30. **Rabin, H., R. H. Neubauer, N. J. Woodside, J. L. Ciemanec, W. C. Wallen, B. A. Lapin, V. A. Agrba, L. A. Yakoleva, and G. N. Chuvirov.** 1976. Virological studies of baboon (*Papio hamadryas*) lymphoma: isolation and characterization of foamy viruses. J. Med. Primatol. **5:**327–337.
31. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313:**277–284.
32. **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.
33. **Riggs, J. L., L. S. Oshiro, D. O. N. Taylor, and E. H. Lennette.** 1969. Syncytium-forming agent isolated from domestic cats. Nature (London) **222:**1190–1191.
34. **Rosen, C. A., J. G. Sodroski, and W. A. Haseltine.** 1985. The location of cis-acting regulatory sequences in the human T cell lymphotropic virus type III (HTLV-III/LAV) long terminal repeat. Cell **41:**813–823.
35. **Schnitzer, T. J.** 1981. Characterization of a simian foamy virus isolated from a spontaneous primate lymphoma. J. Med. Primatol. **10:**312–328.
36. **Schwartz, D. E., R. Tizard, and W. Gilbert.** 1983. Nucleotide sequence of Rous sarcoma virus. Cell **32:**853–869.
37. **Shimotohno, K., Y. Takahashi, N. Shimizu, T. Gojobori, D. W. Golde, I. S. Y. Chen, M. Miwa, and T. Sugimura.** 1985. Complete nucleotide sequence of an infectious clone of human T-cell leukemia virus type II: an open reading frame for the protease gene. Proc. Natl. Acad. Sci. USA **82:**3101–3105.
38. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293:**543–548.
39. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Haase, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42:**369–382.
40. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfizer monkey virus: an immunosuppressive D-type retrovirus. Cell **45:**375–385.
41. **Stancek, D., M. Stancekova, M. Janotka, P. Hnilica, and D. Oravec.** 1975. Isolation and some serological and epidemiological data on the viruses recovered from patients with subacute thyroiditis de Quervain. Med. Microbiol. Immunol. **161:**133–144.
42. **Taylor, J. M.** 1977. An analysis of the role of tRNA species as primers for the transcription into DNA of RNA tumor virus genomes. Biochim. Biophys. Acta **47:**57–71.
43. **Temin, H. M.** 1987. Evolution of retroviruses and other retrotranscripts. UCLA Symp. Mol. Cell. Biol. **71:**1–28.
44. **Thayer, R. M., M. D. Power, M. L. Bryant, M. B. Gardner, P. J. Barr, and P. A. Luciw.** 1987. Sequence relationships of type D retroviruses which cause simian acquired immunodeficiency syndrome. Virology **157:**317–329.
45. **Toh, H., R. Kikuno, H. Hayashida, T. Miyata, W. Kugimiya, S. Inouye, S. Yuki, and K. Saigo.** 1985. Close structural resemblance between putative polymerase of a *Drosophila* transposable genetic element 17.6 and *pol* gene product of Moloney murine leukaemia virus. EMBO J. **4:**1267–1272.
46. **Varmus, H. E.** 1985. Reverse transcriptase rides again. Nature (London) **314:**583–584.
47. **Verma, I. M., N. L. Meuth, and D. Baltimore.** 1972. Covalent linkage between ribonucleic acid primer and deoxyribonucleic acid product of the avian myleoblastosis virus deoxyribonucleic acid polymerase. J. Virol. **10:**622–627.
48. **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.
49. **Weiss, R., N. Teich, H. Varmus, and J. Coffin.** 1982. RNA tumor viruses, p. 25–207. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
50. **Werner, J., and H. Gelderblom.** 1979. Isolation of foamy virus from patients with de Quervain thyroiditis. Lancet **ii:**258–259.
51. **Wong-Staal, F., and R. C. Gallo.** 1985. Human T-lymphotropic retroviruses. Nature (London) **317:**395–403.
52. **Young, D., J. Samuels, and J. K. Clarke.** 1973. A foamy virus of possible human origin isolated in BHK-21 cells. Arch. Gesamte Virusforsch. **42:**228–234.