The EMBO Journal vol.6 no.7 pp.2077–2084, 1987

Nucleotide sequence analysis of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes

Rolf M.Flügel, Axel Rethwilm, Bernd Maurer and Gholamreza Darai[1]

Institute for Virus Research, German Cancer Research Center, Im Neuenheimer Feld 280, and [1]Institute for Medical Virology, University of Heidelberg, Heidelberg, FRG

Communicated by B.Hirt

Recombinant clones that represent the 3′ part of the genome of the human spumaretrovirus (foamy virus) were established from viral DNA and from DNA complementary to viral RNA. The recombinant clones were characterized by blot hybridizations and nucleotide sequence analysis. The deduced protein sequence of the clones at their 5′ ends was found to be homologous to the 3′ domain of retroviral reverse transcriptases. Downstream of a small intergenic *pol*–*env* region a long open reading frame of 985 amino acid residues was identified that according to its genomic location, size, glycosylation signals, and hydrophobicity profile closely resembles the lentiviral *env* genes. The spumaretroviral *env* gene is followed by two open reading frames, termed *bel-1* and *bel-2* which are located between *env* and the long terminal repeat region. The long terminal repeat of 1259 nucleotides is preceded by a polypurine tract and contains the canonical signal sequences characteristic for transcriptional regulation of retroviruses. The provisional classification of the spumaretrovirus subfamily is discussed.

Key words: foamy retrovirus/DNA sequence/reverse transcriptase/transmembrane protein

Introduction

Retroviruses are a family of eukaryotic, single-stranded RNA viruses that replicate through a DNA intermediate, the provirus (Temin, 1976). Retroviruses are the only RNA viruses known to cause cancer and have so far been found in most vertebrates (for review see Weiss *et al.*, 1982). They have been classified into three subfamilies: the onco-, lenti-, and spumaviruses. While members of the oncoviruses and quite recently the lentiviruses have been studied in many laboratories, spumaviruses, also called foamy viruses, have not been characterized in detail (Weiss *et al.*, 1982). Their phylogenetic relatedness to the other subfamilies, and particularly to the human T-cell lymphotropic retroviruses (HTLV-I, HTLV-II) and the human immunodeficiency viruses (HIV), the AIDS viruses, is unknown. In addition, the study of a possible relationship of spumaviruses to human disease has been hampered by their notoriously poor growth properties in tissue cultures. There have been reports (Scolnick *et al.*, 1970; Liu *et al.*, 1977) on the reverse transcriptase from simian foamy virus isolates and on the viral RNA from a human isolate (Loh and Matsuura, 1981). The first human spumaretrovirus (HSRV) was isolated from a patient with a nasopharyngeal carcinoma by Achong *et al.* (1971) and it is this virus strain that was used for this study. In several other instances, spumaviruses were isolated from patients with various diseases; no proven pathogenicity for any of the isolates has been reported (Weiss *et al.*, 1982).

The aim of this work was to gain more insight into the structure and function of foamy viruses. To this end, molecular cloning of viral DNA and cDNA was performed. As a first and fundamental step the resulting recombinant DNA clones were characterized by nucleotide sequence analysis to elucidate the primary structure of the central and the 3′ region of the genome of this interesting virus and to determine whether or not novel genes are encoded by the HSRV genome as well as the anticipated retroviral genes. Another objective of this study was to establish molecular clones specific for HSRV for screening DNAs from human patients who suffer from diseases that are suspected to have an unknown viral etiology.

Results

Identification of molecular clones containing HSRV-specific sequences

Two approaches were used for the molecular cloning of HSRV-specific DNA. In the first, recombinant plasmids were constructed from viral DNA using genomic DNA of HSRV-infected human embryonic lung (HEL) cells that was digested with *Bam*HI. The resulting *Bam*HI DNA fragments were inserted into the *Bam*HI site of the plasmid vector pAT153 (Twigg and Sheratt, 1980). One of the resulting recombinant clones termed pHSRV-B52 (B52) hybridized to HSRV cDNA and was characterized in detail. As a second approach, recombinant λ phage clones were established using the *Hin*dIII digests of cDNA that had been synthesized from virions of HSRV. One of the resulting recombinant λ clones was subcloned into the *Hin*dIII site of pAT153 and termed pHSRV-H-C55 (C55).

The 5.4-kbp insert of the recombinant plasmid C55 hybridized to the ^{32}P-labeled insert of B52 DNA of 2.2 kbp and vice versa (Figure 1, lanes 9–11). Restriction enzyme analysis unambiguously demonstrated that the viral inserts share a common region of 817 bp that is located at the 3′ end of the C55 insert and at the 5′ end of the B52 insert as marked by arrows in Figure 1B and is also shown in Figure 2.

To demonstrate that those recombinant plasmids that had been established from viral DNA and that hybridized to cDNA in fact contained sequences specific for HSRV, blot hybridization experiments were performed using B52 and C55 DNAs and human genomic DNAs from uninfected and HSRV-infected HEL cells. The results shown in Figure 1 (A–E, lanes 1–8) clearly indicate that only DNAs from HSRV-infected cells gave positive hybridization signals (lanes 5–8), whereas DNA from uninfected cells (lanes 1–4) were negative indicating that the inserts of B52 and C55 have specific DNA sequences homologous to sequences of infected cells as expected e.g. for an infection by an exogenous retrovirus.

The results of the restriction analysis of viral DNA sequences of both recombinant plasmids are in complete agreement with the physical maps shown in Figure 2 and clearly demonstrated that the 3′ *Hin*dIII site of clone C55 is located 817 nucleotides downstream of the *Bam*HI site at the C55/B52 boundary (Figures 1 and 2).

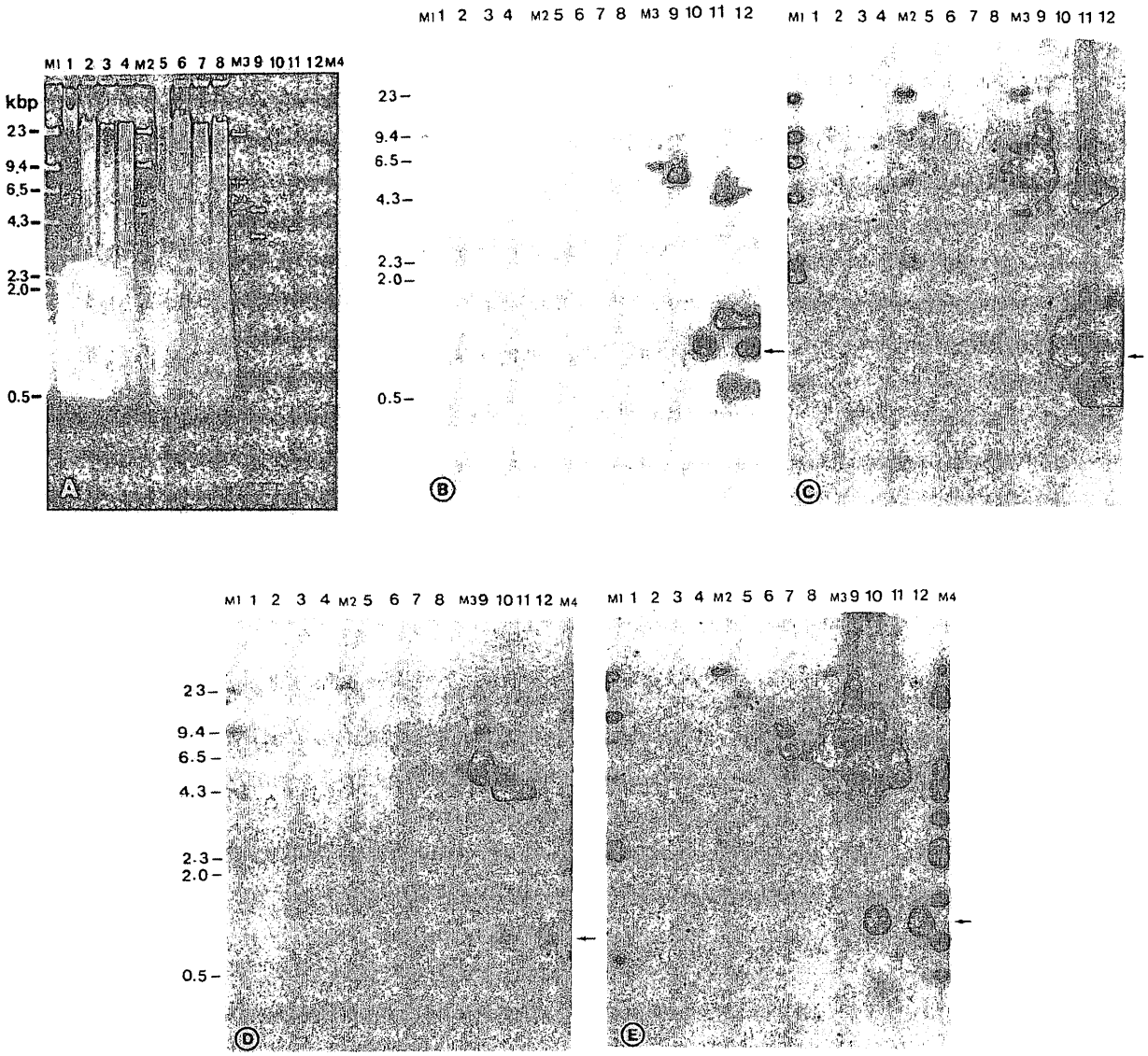


Fig. 1. Southern blot hybridizations of HSRV-specific DNA derived from recombinant plasmids pHSRV-B52 and pHSRV-H-C55 to each other and to DNAs from uninfected and HSRV-infected HEL cells. DNAs (5 μg) from uninfected **(lanes 1–4)** and from HSRV-infected HEL cells **lanes 5–8)** were analyzed undigested (lanes 1 and 5); digested with *Bam*HI (lanes 2 and 6), *Hin*dIII (lanes 4 and 8) and double-digested with *Bam*HI/*Hin*dIII (lanes 3 and 7). DNAs of pHSRV-H-C55 **(lanes 9** and **10)** and of pHSRV-B52 **(lanes 11** and **12)** were cleaved with *Hin*dIII (9 and 11) and double-digested with *Bam*HI/*Hin*dIII (lanes 10 and 12). DNAs were separated electrophoretically on a 0.8% agarose gel. Unlabeled and ^{32}P-labeled λ DNA digested with *Hin*dIII (M1); *Mlu*I (M2), *Eco*RI (M3), and *Cla*I (M4) served as markers and a control for electrophoretic transfer of DNA fragments to nitrocellulose filter. **A**, ethidium bromide staining; **B** and **C**, autoradiograph of the same gel after hybridization to ^{32}P-labeled insert of recombinant pHSRV-B52; and in **D** and **E** to the ^{32}P-labeled insert of recombinant pHSRV-C55 at different times of exposure (B and D, 20 h; C and E, 72 h). Arrows mark the position of the 816-bp hybridizing DNA fragments of the region common to both recombinants.

Nucleotide sequence analysis of pHSRV-H-C55 and pHSRV-H-B52

The strategy used for determining the primary structures of B52 and C55 DNA and restriction maps for some enzymes are given in Figure 2. More than 95% of the DNA sequence was determined for both strands and various multiple-cut restriction enzymes in addition to those shown in Figure 2 were used to verify the sequence, in particular the enzymes *Taq*I, *Bst*EII, *Mae*II, *Dde*I, *Hin*fI, and *Sau*3A I were used for sequencing in addition to those shown in Figure 2. The sequence was determined for B52 DNA (from nucleotide position 4549 to 6755, Figure 3) at which it terminates. Since restriction fine mapping had revealed that recombinant clone C55 is co-linear with clone B52, overlaps it and extends the B52 sequence into the 5′ direction (Figures 2 and 3), the 5′ part of the sequence (position 1–4549) shown in Figure 3 was derived from C55 DNA. The region common to both viral inserts comprises 817 bp (from the *Bam*HI at 4548 to the *Hin*dIII site at 5365, Figure 3).

The resulting nucleotide sequence of the viral DNA inserts of both recombinants are shown in Figure 3 with the predicted amino acid sequences of the corresponding HSRV gene products. Open reading frames longer than 129 amino acid residues were not present on the opposite strand, consistent with other retroviruses.

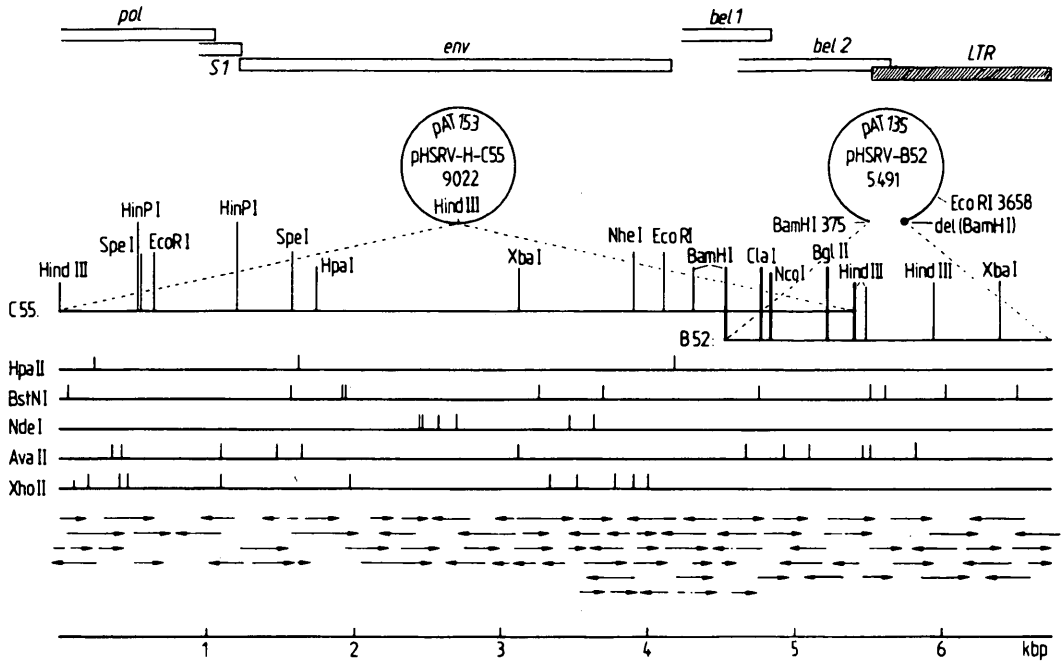


Fig. 2. Restriction maps of DNA viral inserts of two recombinant plasmids (pHSRV-H-C55 and pHSRV-B52) harboring HSRV-specific sequences and the strategy for determining the nucleotide sequence. The upper part shows the viral genes (*pol, S1, env, bel 1* and *2* and LTR) corresponding to the inserts of two recombinant plasmids C55 and B52 that overlap in the *bel* region. Restriction maps of both viral inserts are given for some of those enzymes used for sequencing. The arrows below the maps indicate the direction and extent of sequences determined for each fragment. The lower line represents the scale in kbp. The filled circle in the recombinant B52 marks the *Bam*HI site of pAT153 that was partially deleted, probably during molecular cloning of unintegrated linear HSRV DNA. This assumption is supported by the nucleotide sequence analysis of B52 DNA which indicated that the 3′ terminus of B52 DNA coincided with the 3′ end of the viral DNA genome (see Figure 3).

The nucleotide sequence presented in Figure 3 includes the 3′ domain of the *pol* gene, the entire *env* of HSRV along with its 5′ and 3′ flanking regions. The main features of the sequence are presented starting from the 5′ to the 3′ end of the viral genome.

3′-pol endonuclease (integrase) sequences

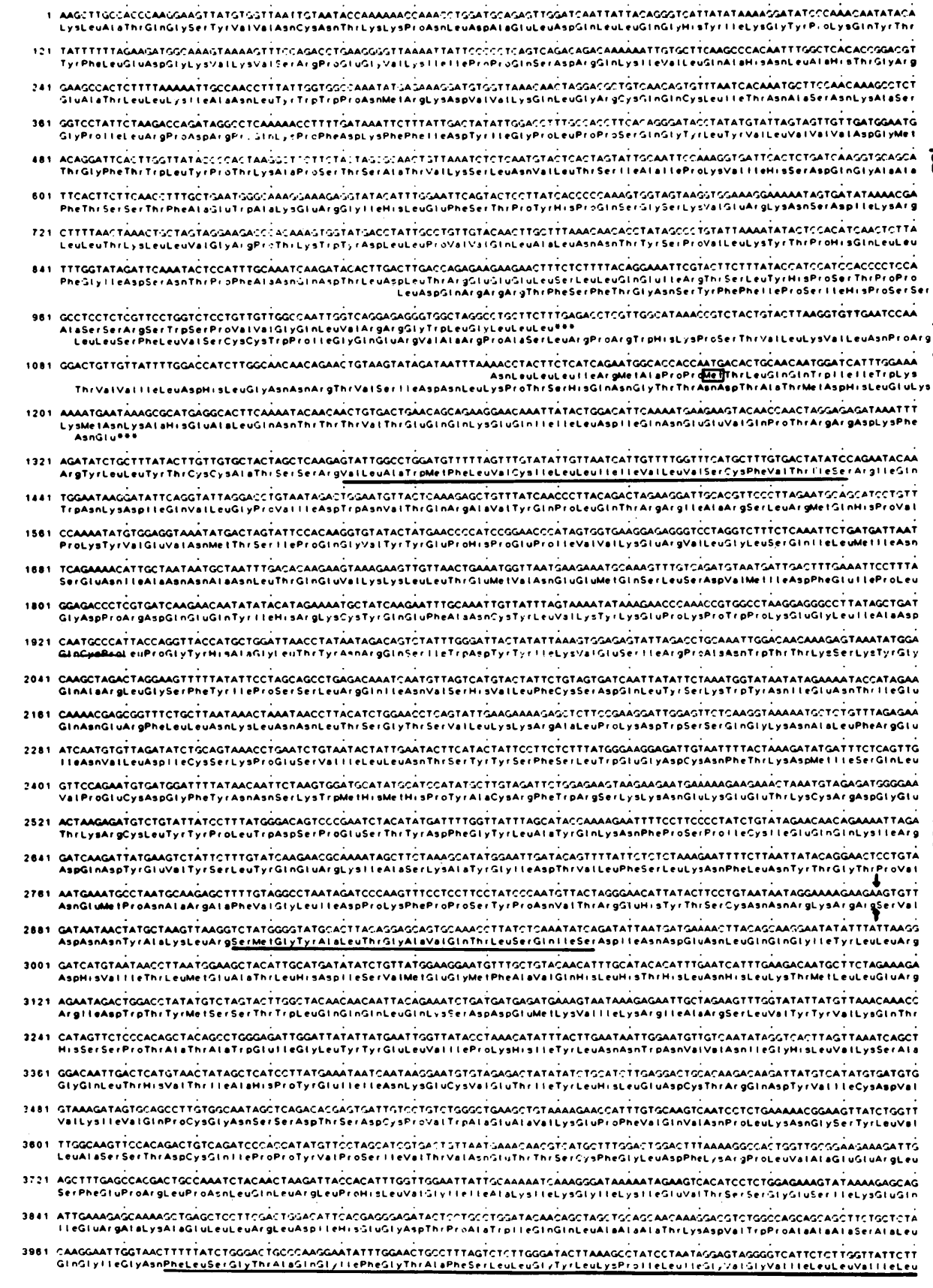
Starting from the *Hin*dIII site (nucleotides 1–6 in Figure 3), an uninterrupted reading frame runs to nucleotide 1030. The corresponding protein sequence is homologous to the 3′ domain of the reverse transcriptases from other retroviruses. Counting identical amino acid residues, degrees of homologies range from 22 and 23% to 30.5% when compared to the 3′ domains of VIV, HIV and Mo-MLV reverse transcriptases (Shinnick *et al.*, 1981; Sonigo *et al.*, 1985; Ratner *et al.*, 1985; Wain-Hobson *et al.*, 1985). At the nucleotide level, the degree of homology was found to be 46.1, 46.5 and 48.0% in comparison to VIV, MLV and HIV. The direct comparison between the corresponding retroviral proteins and the HSRV sequence shown in Figure 4 reveals that virtually all amino acid residues conserved in HIV, VIV, and MLV are found to be invariant in this part of the HSRV *pol* gene product (identities are marked by asterisks). If similar amino acid residues are taken into account an even higher homology of 47.3–55.2% is reached for HIV and MLV, respectively.

pol–env intergenic region

Computer analysis of the carboxy terminal region of the *pol* gene identifies a short reading frame of a hypothetical HSRV protein that starts at nucleotide position 888 and continues for 107 amino acid residues (Figure 3) in a reading frame different from that of *pol*. In analogy to the EIV genome that potentially codes for a short protein in the *pol–env* intergenic region (Rushlow *et al.*, 1986), the corresponding HSRV gene product is designated as *S1*. This hypothetical protein would be required to be synthesized from a spliced RNA. *S1* of HSRV does not show any homology to the *S1* of EIV, nor to any other retroviral proteins. Its sequence has three glycosylation signals of the type Asn-X-(Ser/Thr). The 5′ part of the HSRV *S1* overlaps the COOH-terminal region of the HSRV *pol* gene by 141 nucleotides and its termination codon is located at nucleotide 1209, thus overlapping the HSRV *env* region that lies in a different frame when compared to either *S1* or *pol* (Figure 3).

env gene

The major open reading frame of C55 DNA begins at nucleotide 1141 and extends for 995 codons, terminating with TAG at 4126 (Figure 3). The second start codon ATG for *env* is located 11 triplets in, and its flanking sequences conform to those of a typical initiator (Kozak, 1984). Thus, the HSRV *env* precursor would have a mol. wt of 113 kd, very close to that of the VIV *env* precursor reported to be 115 kd (Sonigo *et al.*, 1985). This unusually long open reading frame clearly is the viral envelope gene, since its protein sequence has three hydrophobic regions, 14 potential N-linked glycosylation sites of the type Asn-X-Ser/Thr, and last but not least a proteolytic cleavage signal Arg-Lys-Arg-Arg typical of retroviral *env* protein sequences (Seiki *et al.*, 1983). Closer examination of the HSRV *env* protein sequences shows that the first 62 amino acid residues after the initiator Met at 1171 are predominantly hydrophilic in character resembling in that respect the *env* precursors of VIV and mouse mammary tumor virus (MMTV) (Majors and Varmus, 1983; Redmond and Dickson, 1983; Sonigo *et al.*, 1985). The first


pol
env
R1
R3

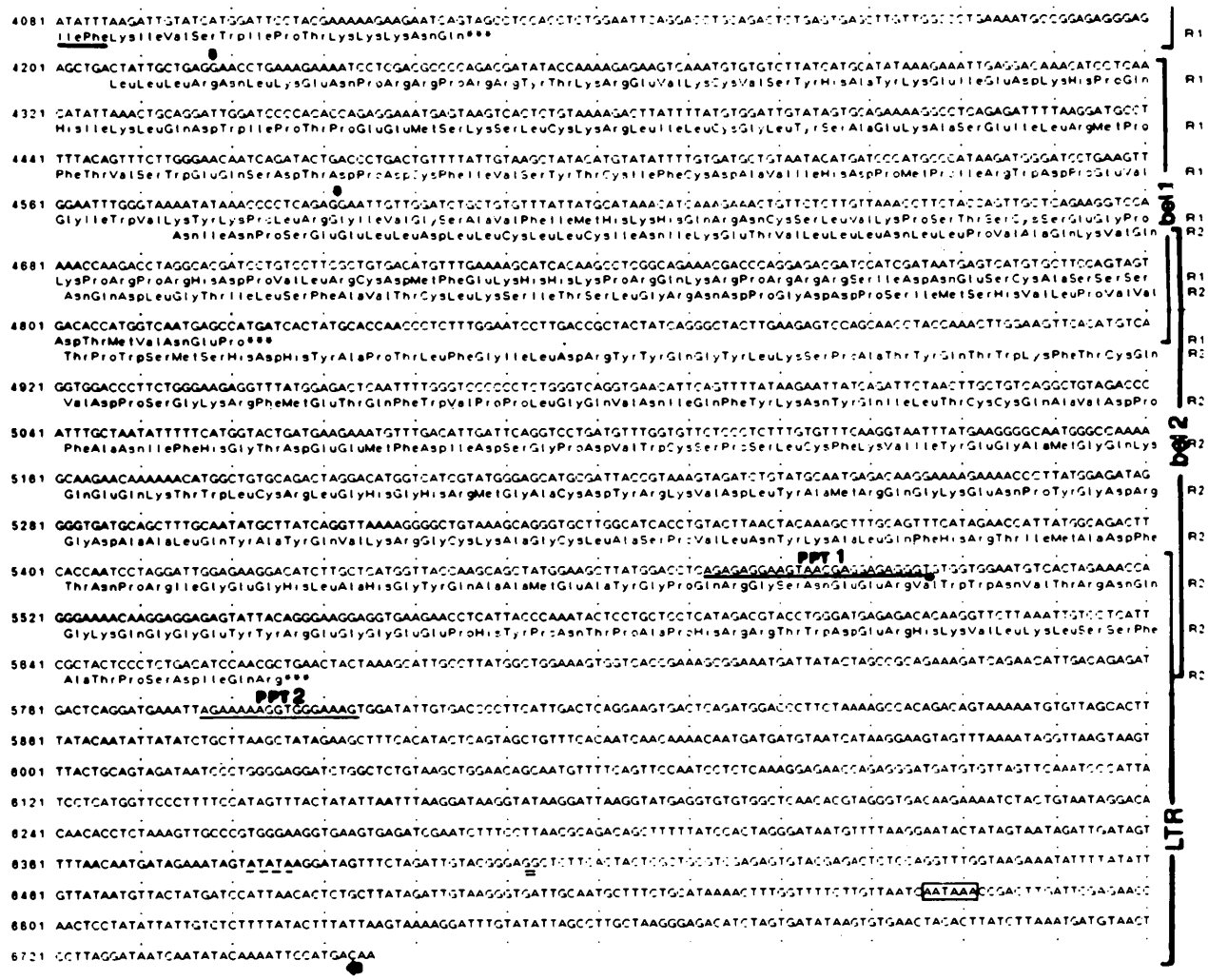


Fig. 3. DNA sequences of 6755 bp of viral inserts of pHSRV-H-C55 and B52. The predicted amino acid sequences encoded by the *pol*, *S1*, *env*, *bel 1*, and *bel 2* genes of HSRV are shown below the DNA sequence. One of the potential start codons for the *env* protein is boxed, and three of its hydrophobic regions corresponding to the signal peptide, the fusion sequence, and the transmembrane sequence are underlined. Stop codons are marked by three asterisks. The proteolytic cleavage signal for generating the outer membrane and the transmembrane protein of *env* is indicated by vertical arrows above and below the cleavage site. Small thick vertical arrows mark potential splice acceptor sites for the *bel* genes. Horizontal small thick arrows indicate the inverted repeats at the start and end of the LTR. The polypurine tracts (PPT) are underlined, a potential poly(A) addition signal is boxed, the putative cap site is doubly underlined and the TATA box is marked by a broken line.

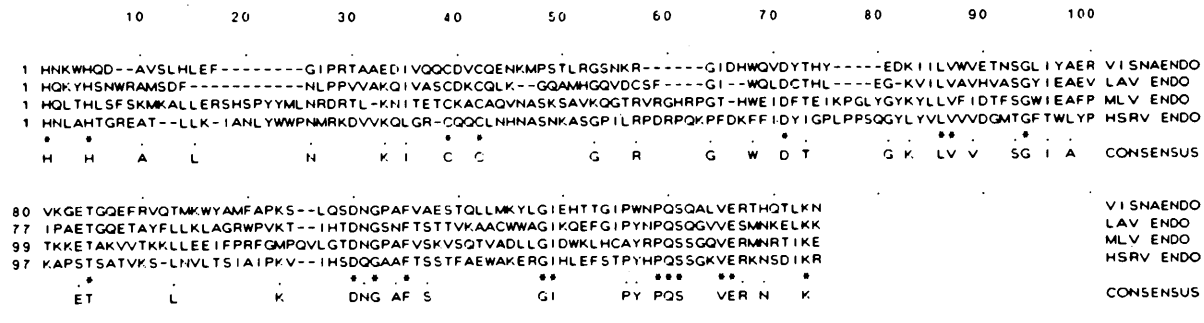


Fig. 4. Region of homologous protein sequences of the endonuclease domains of the reverse transcriptase of VIV, HIV (LAV-isolate), Mo-MLV, and HSRV. Amino acid residues identical for all four retroviral reverse transcriptase domains are marked by asterisks, and by points when three out of four residues are identical. Gaps were introduced to maximize homology. The one-letter code for abbreviating amino acids is used.

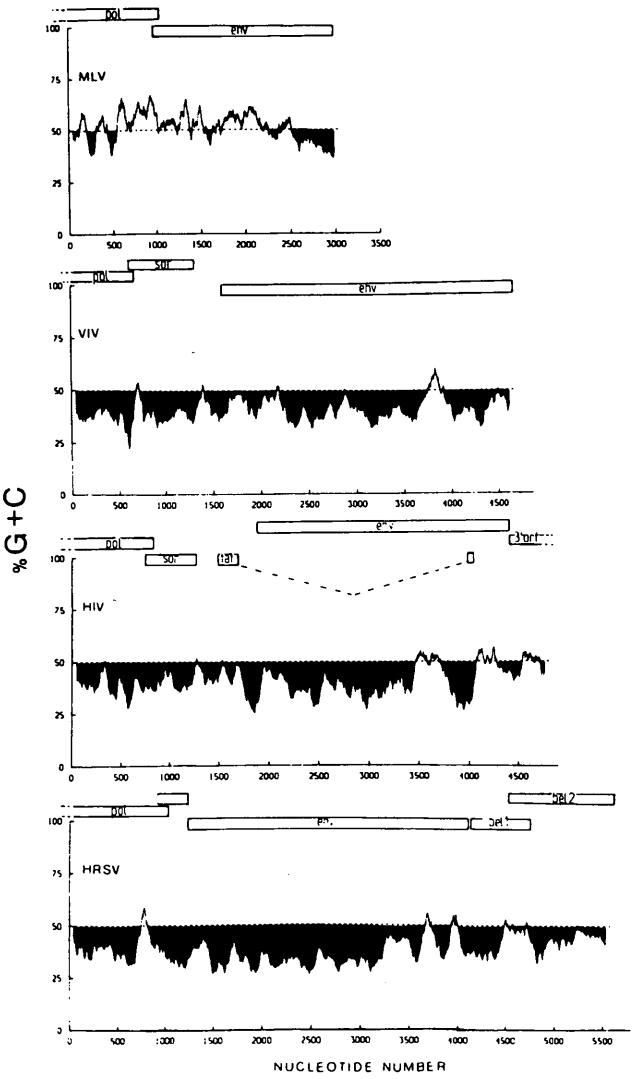


Fig. 5. Base composition of coding strands of MLV, VIV, HIV, and HSRV 3′ *pol*–*env* regions. The (G + C) molar content was computed by using a window size of 100 bases. All sequences begin with the endonuclease domain of the *pol* gene and end at the 3′ LTR.

hydrophobic region of the HSRV *env* is identified by the hydropathic profile (Kyte and Doolittle, 1982) and comprises 24 amino acids (residues 64–87, Figure 3) that presumably is the signal peptide. A second less hydrophobic region of the HSRV *env* of 17 residues is preceded by the typically basic sequence Arg-Lys-Arg-Arg. Similar sequences occur in most retroviral *env* precursor molecules and represent the sites of cleavage by a cellular protease that give rise to an outer membrane (OMP) (or external glycoprotein) and a carboxy-terminal transmembrane protein (TMP) (Seiki *et al.*, 1983; Coffin, 1986). The third hydrophobic region of 36 residues probably corresponds to the transmembrane domain (Figure 3). It is interrupted by one Lys residue as is the case in the corresponding domains of the HIV *env* (Ratner *et al.*, 1985; Wain-Hobson *et al.*, 1985).

The mol. wts of foamy viral glycoprotein precursors are unknown. The single report by Benzair *et al.* (1985) of a major *env* glycoprotein of simian foamy virus type 1 of 70 kd cannot directly be compared to the OMP of the HSRV, since the unmodified protein sequence predicts a value of 53.7 kd. However, the value of 53.7 kd is consistent with the reported one, since it is a general characteristic for most retrovirus that the mol. wts calculated from gene sequences are lower when compared to the values determined by SDS–PAGE (Rushlow *et al.*, 1986). The relatively high apparent mol. wts for retroviral glycoproteins analyzed by SDS–PAGE have been explained by the effects of protein glycosylation on migration in polyacrylamide gels (Rushlow *et al.*, 1986). This holds also true for the mature OMP of HSRV that contains 12 potential glycosylation signals which would account for the difference in mol. wts neglecting that a spumavirus from a different species was used in that report (Benzair *et al.*, 1985).

Most of the glycosylation sites are located in the NH_2-terminal half of the *env* sequence (Figure 3). This pattern agrees well with that of HIV (Ratner *et al.*, 1985; Wain-Hobson, 1985), EIV (Rushlow *et al.*, 1986), and VIV (Sonigo *et al.*, 1985) and is consistent with models for retroviral *env* structures and its interaction with the cytoplasmic membrane in which the OMP is the externally exposed major glycoprotein whereas TMP, as the smaller *env* protein, forms the more sequestered transmembrane component (Bolognesi *et al.*, 1978; Lenz *et al.*, 1982; Seiki *et al.*, 1983). At the nucleotide level, the degree of homology of the HSRV *env* gene is highest to the VIV gene with 52.1 and 48.7% to HIV (LAV). For comparison, the corresponding match for HIV and VIV is 51.4%, close to the value reported above.

Sequences between env and the 3′ LTR (bel 1 and 2)

Analysis of the post-*env* region of HSRV identifies two open reading frames, termed *bel 1* and *bel 2* which are 205 and 364 amino acid residues long (Figure 3). Probably, both putative proteins coded by these open reading frames would have to be generated from spliced viral RNAs. Candidates for the corresponding splice acceptor sites are marked by vertical arrows in Figure 3. Particularly, the second one conforms well to the consensus splice acceptor signals that are also found in the HTLV genome for mRNAs from which the *px* proteins are derived and in which the corresponding initiator codon is located on a RNA leader sequence that is derived from the 5′ part of the genome (Shimotohno *et al.*, 1984). Alternatively, shorter versions of the *bel* proteins might be initiated from the corresponding first initiator codons (Figure 3). Striking homologies to known proteins recorded in the NBRF-PIR protein sequence data base were not found. However, the overall structure of the *bel* proteins are quite interesting. *Bel 1* has 10 Cys residues, 2 glycosylation signals, and a strongly basic region of 13 residues proximal to its COOH terminus, whereas *bel 2* has 12 Cys residues, one glycosylation signal, and an extremely hydrophilic stretch of amino acid residues proximal to its COOH terminus.

The LTR region

The 3′ last 1259 nucleotides of the B52 DNA sequence contain a number of key features of regulatory signals shown to be required for retroviral replication and transcription. There is a characteristic sequence, AGAGAGGAAGTAACGAGGAGAGGG, the polypurine tract (Figure 3), that defines the 5′ boundary of the HSRV LTR and is found in all retroviral genomes at this location and is the putative primer for plus strand DNA synthesis of retroviruses. Within the 3′ presumed LTR there is one poly(A) addition signal AATAAA (position 6576–6581) which perfectly matches the canonical consensus sequence. If the location of the poly(A) signal is assumed to be non-variant with respect to the polyadenylation site of HSRV, one can draw an analogy to the

LTRs of HIV and VIV (Ratner *et al.*, 1985; Sonigo *et al.*, 1985) and place the polyadenylation site 24 bp downstream of the poly(A) addition signal at CA (position 6600). Similar arguments indicate that the cap site starts with a G close to position 6412. A perfect and multiple TATATA box precedes the presumptive RNA initiation site by 31 bp. Thus, the HSRV LTR can be subdivided into three regions. The R and U5 regions were calculated to be 189 and 154 bp, respectively. The U3 region was estimated to be 916 bp long. Restriction fine mapping of another recombinant clone that was derived from viral DNA and that contained sequences from the 5′ region of the HSRV genome resulted in DNA fragment sizes that were consistent with those obtained with the 3′ LTR sequences from B52 DNA (R.M.Flügel, unpublished observation).

Although the HSRV LTR is unusually long, open reading frames for encoding proteins of >70 amino acid residues were not found. Furthermore, the HSRV LTR had little if any sequence homology to the LTRs of other retroviruses, in particular to those of HIV-I, HTLV-II, VIV, MMTV, and HIV-2 (Fasel *et al.*, 1982; Kennedy *et al.*, 1982; Ratner *et al.*, 1985; Shimotohno *et al.*, 1985; Sonigo *et al.*, 1985; Clavel *et al.*, 1986). In addition, no significant sequence homology was found to human endogenous retrovirus-like sequences including the 968-bp long LTR sequence of human endogenous retroviral HERV genes (Ono, 1986).

There are 12 octamer sequences, six nonamers and one decamer that form direct repeats within the LTR. The indirect repeats include a decamer at positions 6646 and 6743 with a loop size of 78 bp and two octamers with a relatively small loop size of 34 and 5 bp, respectively. The significance of these repeats with respect to the regulation of transcription remains to be elucidated.

A comparison of the base composition of the HSRV-coding strand in the 3′ *pol*−*env* region with those of three other retroviruses, namely HIV, visna virus (VIV), and murine leukemia virus (Mo-MLV) is presented in Figure 5. In the 3′ *pol*−*env* region the coding strands of the lentiviruses VIV and HIV, and of HSRV have a very high adenine + thymidine content of >60−62 mole percent and HSRV having 61.7 percent which contrasts with that of oncoviruses, MLV has a higher G + C percent as shown for comparison in the upper panel of Figure 5. Thus, the HSRV 3′ *pol*−*env* region shares a low G + C content with all other lentiviruses, including equine infectious anemia virus (EIV) (Stephens *et al.*, 1986; Rushlow *et al.*, 1986).

Discussion

The organization of the coding sequences of the central and 3′ half of the HSRV genome is 5′ *pol* (endonuclease)−*env*−*bel 1*−*bel 2*−LTR 3′ as judged by protein sequence homology with the endonuclese domains of other retroviral reverse transcriptases. The 3′ *pol* domain of HSRV is homologous to the corresponding regions of retroviral reverse transcriptases. Counting identities only, the degree of protein homology is 22% for HIV. It is interesting that this homology is 30% for the MLV *pol* endonuclease domain. However, it is remarkable that in a comparison of this HSRV sequence with those of two lentiviruses HIV and VIV and one oncovirus, MLV, those amino acid residues well known to be conserved can be aligned so that these residues are invariant, including two of each equidistantly spaced His and Cys residues, forming a 'finger', as postulated by Johnson *et al.* (1986). This homology exemplifies the close structural relatedness between this part of the HSRV genome and the corresponding genes of other retroviruses. The significance of this homology is underscored when similar amino acid residues are taken into account which results in a degree of homology of 47−55%.

The reading frame of the HSRV *pol* gene does not overlap that of the *env* gene and in this respect HSRV appears to resemble the lentiviruses, HIV, VIV, and EIV, but is clearly different from the oncoviruses, the D- and B-type viruses (Sonigo *et al.*, 1986), and the HTLV group all of which have overlapping *pol*/*env* genes (Weiss *et al.*, 1985). It is of interest that the HSRV intergenic region does not seem to have a gene equivalent to either the *sor* genes of HIV or VIV (Arya and Gallo, 1986). Instead HSRV encodes a small gene, *S1*, that partially overlaps the carboxy-terminal part of the *pol* and the amino-terminal end of *env*, whereas the remainder of *S1* is encoded by the HSRV *pol*−*env* intergenic region reminiscent of the *S1* of EIV (Rushlow *et al.*, 1986). The exceptionally long *env* gene of HSRV of at least 985 residues has a counterpart in the VIV *env* reported to consist of 983 amino acid residues, not taking the stop codon within the TMP of the VIV *env* into account (Sonigo *et al.*, 1985). Unlike most retroviruses but like lentiviruses, the TMP of the HSRV *env* gene does not contain the immunosuppressive domain of Cianciolo (1985; Sonigo *et al.*, 1986). While the overall structural similarities between the *env* genes of HSRV and VIV are striking, the post-*env* region of the HSRV genome is completely different in size and sequence from that of the VIV genome.

Two novel genes, *bel 1* and *bel 2*, of 205 and 364 amino acid residues are presumably derived from spliced HSRV mRNAs. The *bel* protein sequences are rich in cysteine residues, possess glycosylation signals of the type Asn-X-Thr/Ser, and have strongly basic stretches, so that they could be compared to growth factors, hormone receptors and other components of the cellular signal chain. Alternatively, they might function as trans-acting transcriptional activators as recently reported for HIV (Sodroski *et al.*, 1985; Rosen *et al.*, 1986) and HTLV-I (Fujisawa *et al.*, 1985).

Since there is no obvious homology of the *bel* sequences to any of the known retroviral sequences including the 3′ *orf* of HIV, the *tat* genes of HTLV-I, II, and BLV, and the *orf* of the MMTV LTR, it seems that HSRV is different from previously well characterized retroviruses. This conclusion seems to be reinforced by the finding that the homology of the HSRV 3′ *pol* domain to that of MLV is somewhat greater than to those of the lentiviruses. Furthermore, the region between the HSRV *env* and the 3′ LTR is of extraordinary length when compared to all other retroviruses. On the other hand, the fact that the primer binding site of HSRV closely resembled that of VIV, HIV and of caprine arthritis encephalitis virus (Sherman *et al.*, 1986) (tRNA-lys, our unpublished data) clearly indicates that HSRV is closely related to the lentivirus subfamily. The relatively high adenine plus thymine content of the central and 3′ part of the HSRV genome is strikingly similar to that of lentiviruses. Taken together, one can conclude that HSRV is a member of a retrovirus subfamily that is different from oncoviruses, but also different from other groups such as type D, type B retroviruses and the HLTV/BLV group. However, a final and definite phylogenetic placement has to take the 5′ *pol* and *gag* sequences of the HSRV genome into account.

Of potentially great interest are the *bel* genes, that are located in a genomic region in which either viral oncogenes or *tat* genes have been found. Further analyses of these genes and their

putative gene products have to be performed to learn more about their functions.

Materials and methods

Cells and virus

Cells of human embryonic lung fibroblasts were prepared as described previously (Flügel *et al.*, 1987). Virus was kindly provided by Dr P.Loh and infection of HEL cells was carried out as described (Loh and Matsuura, 1981).

Construction of recombinant plasmids

As source for viral DNA, total DNA from HSRV-infected HEL cells was extracted, deproteinized, and run on a 0.8% low-melting agarose gel. After staining with ethidium bromide, DNA bands were divided into five fractions (A–E) and isolated from agarose. Aliquots of the resulting DNA bands were rerun and visualized by ethidium bromide staining. The discrete and highly intense DNA bands of fraction B were isolated and digested with various restriction enzymes, including *Bam*HI. A *Bam*HI RNA fragment of 2.2 kbp highly enriched for viral DNA was isolated and used for molecular cloning. Recombinant plasmids containing this *Bam*HI DNA fragment were constructed using the plasmid vector pAT153 (Twigg and Sheratt, 1980). Recombinant λ clones were established from cDNA by inserting the *Hin*dIII digest of cDNA into the corresponding sites of the λ vector NM1149 (Murray, 1983). The recombinant λ clones were subcloned into the plasmid vector pAT153. Selection, amplification, and purification of the recombinants were carried out as described previously (Koch *et al.*, 1977). The molecular cloning of the HSRV genome is the subject of a manuscript in preparation (A.Rethwilm *et al.*).

Preparation of cDNA

HSRV was grown on HEL cells and purified by sucrose gradient centrifugation. For synthesis of cDNA for molecular cloning, the procedure of Rothenberg and Baltimore (1976) was followed for the first strand, and the method of Gubler and Hoffmann (1983) was used for the second strand. Alternatively, HSRV from clarified HEL cell supernatants was purified by two sucrose gradient centrifugations. Synthesis of cDNA for hybridization probes was carried out according to Yoshida *et al.* (1982) using oligodeoxynucleotide primers provided by Dr John Taylor, Fox Chase Cancer Center, Philadelphia (Taylor *et al.*, 1976).

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]dNTP defined by the recognition sequence of the given endonuclease. Alternatively, T4 polynucleotide kinase and [γ-^{32}P]ATP were used for labeling. The labeled fragments were sequenced according to the method of Maxam and Gilbert (1977). More than 95% of the sequence was done for both strands. Analysis of the sequence data was performed using the BSA program deviced by Dr S.Suhai at the German Cancer Research Center, Heidelberg.

Nucleic acid hybridization

DNA was cleaved by different restriction enzymes, separated by agarose slab gel electrophoresis, electrophoretically transferred to nitrocellulose paper, and hybridized according to Southern (1975). Aliquots (1 μg) of individual DNAs were labeled *in vitro* as described by Rigby *et al.* (1977). Each sample (50 μl) contained 100 μCi of [α-^{32}P]dATP and [α-^{32}P]dCTP (3000–5000 Ci/mmol).

Materials

Labeled compounds were purchased from Amersham and New England Nuclear. Enzymes from Boehringer or from New England Biolabs.

Acknowledgements

We thank Howard Temin for critically reading the manuscript and Harald zur Hausen for initiating and supporting this project. The excellent assistance of Helmut Bannert is highly appreciated. This project is supported by a postdoctoral training stipend (627-1-1) from the Deutsche Forschungsgemeinschaft.

References

Achong,B.G., Mansell,P.W., Epstein,M.A. and Clifford,P. (1971) *J. Natl. Cancer Inst.*, **46**, 299–307.
Arya,S.K. and Gallo,R.C. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 2209–2213.
Benzair,A.-B., Rhodes-Feuillette,A., Lasneret,J., Emanoil-Ravier,R. and Périès,J. (1985) *J. Gen. Virol.*, **66**, 1449–1455.
Bolognesi,D.P., Montelaro,R.C., Frank,H. and Schäfer,W. (1978) *Science*, **199**, 183–186.
Cianciolo,G.J., Copeland,T.D., Oroszlan,S. and Snyderman,R. (1985) *Science*, **230**, 453–455.
Clavel,F., Guyader,M., Guétard,D., Sallé,M., Montagnier,L. and Alizon,M. (1986) *Nature*, **324**, 691–695.
Coffin,J.M. (1986) *Cell*, **46**, 1–4.
Fasel,N., Pearson,K., Buetti,E. and Diggelmann,H. (1982) *EMBO J.*, **1**, 3–7.
Flügel,R.M., Maurer,B., Bannert,H., Rethwilm,A., Schnitzler,P. and Darai,G. (1987) *Mol. Cell Biol.*, **7**, 231–236.
Fujisawa,J.I., Seiki,M., Kikokawa,T. and Yoshida,M. (1985) *Proc. Natl. Acad. Sci. USA*, **81**, 6349–6353.
Gubler,U. and Hoffman,B.J. (1983) *Gene*, **25**, 263–269.
Johnson,M.S., McClure,M.A., Feng,D.-F., Gray,J. and Doolittle,R.F. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 7648–7652.
Kennedy,N., Knedlitschek,G., Groner,B., Hynes,N.E., Herrlich,P., Michalides,R. and van Ooyen,A.J.J. (1982) *Nature*, **295**, 622–624.
Koch,H.-G., Delius,H., Matz,B., Flügel,R.M., Clarke,J. and Darai,G. (1977) *J. Virol.*, **55**, 86–95.
Kozak,M. (1984) *Nucleic Acids Res.*, **12**, 857–872.
Kyte,J. and Doolittle,R.F. (1982) *J. Mol. Biol.*, **157**, 105–132.
Lenz,J., Crowther,R., Straceski,A. and Haseltine,W. (1982) *J. Virol.*, **42**, 519–529.
Liu,W.T., Natori,T., Chang,K.S.S. and Wu,A.M. (1977) *Arch. Virol.*, **55**, 187–200.
Loh,P.C. and Matsuura,F.S. (1981) *Arch. Virol.*, **68**, 53–58.
Majors,J.E. and Varmus,H.E. (1983) *J. Virol.*, **47**, 495–504.
Maxam,A. and Gilbert,W. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 560–564.
Murray,N. (1983) In Hendrix,P.W., Roberts,J.W., Stahl,F.W. and Weisberg,P.A. (eds), *Lambda II.* Cold Spring Harbor Laboratory, NY, chapter 18, pp. 395–432.
Ono,M. (1986) *J. Virol.*, **58**, 937–944.
Ratner,L., Haseltine,W., Patarca,R., Livak,K.J., Starcich,B., Josephs,S.F., Doran,E.R., Rafalski,J.A., Whitehorn,E.A., Baumeister,K., Ivanoff,L., Petteway,S.R.,Jr, Pearson,M.L., Lautenberger,J.A., Papas,T.S., Ghrayeb,J., Chang,N.T., Gallo,R.C. and Wong-Staal,F. (1985) *Nature*, **313**, 277–284.
Redmond,S.M.S. and Dickson,C. (1983) *EMBO J.*, **2**, 125–131.
Rigby,P.W.J., Dieckmann,M., Rhodes,C. and Berg,P. (1977) *J. Mol. Biol.*, **114**, 237–256.
Rosen,C.A., Sodroski,J.G., Campbell,K. and Haseltine,W.A. (1986) *J. Virol.*, **57**, 379–384.
Rothenberg,E. and Baltimore,D. (1976) *J. Virol.*, **17**, 168–174.
Rushlow,K., Olsen,K., Steigler,G., Payne,S.L., Montelaro,R.C. and Issel,C.J. (1986) *Virology*, **155**, 309–321.
Scolnick,E., Rands,E., Aaronson,S.A. and Todaro,G.J. (1970) *Proc. Natl. Acad. Sci. USA*, **67**, 1789–1796.
Seiki,M., Hattori,S., Hirayama,Y. and Yoshida,M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3618–3622.
Sherman,L., Gazit,A., Yamir,A., Dahlberg,J.E. and Tronick,S.R. (1986) *Virus Res.*, **5**, 145–155.
Shimotohno,K., Wachsman,W., Takahashi,Y., Golde,D.W., Miwa,M., Sugimura,T. and Chen,I.S.Y. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 6657–6661.
Shinnick,T.M., Lerner,R.A. and Sutcliffe,J.G. (1981) *Nature*, **293**, 543–548.
Sodroski,J., Rosen,C., Wong-Staal,F., Salahuddin,S.Z., Popovic,M., Arya,S.K., Gallo,R.C. and Haseltine,W.A. (1985) *Science*, **227**, 171–173.
Sonigo,P., Alizon,M., Staskus,K., Klatzmann,C., Cole,S., Danos,O., Retzel,E., Tiollais,P., Haase,A. and Wain-Hobson,S. (1985) *Cell*, **42**, 369–382.
Sonigo,P., Barker,C., Hunter,E. and Wain-Hobson,S. (1986) *Cell*, **45**, 375–385.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Stephens,R.M., Casey,J.W. and Rice,N.R. (1986) *Science*, **231**, 589–594.
Taylor,J., Ilmensee,R. and Summers,J. (1976) *Biochim. Biophys. Acta*, **442**, 324–330.
Temin,H.M. (1976) *Science*, **192**, 1075–1080.
Twigg,A. and Sheratt,D. (1980) *Nature*, **283**, 216–218.
Wain-Hobson,S., Sonigo,P., Danos,O., Cole,S. and Alizon,M. (1985) *Cell*, **40**, 9–17.
Weiss,R., Teich,N., Varmus,H. and Coffin,J. (eds) (1982) *RNA Tumor Viruses.* Cold Spring Harbor Laboratory, NY, 2nd edition, chapter 2, pp. 25–207.
Yoshida,M., Miyoshi,I. and Himuma,Y. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 2031–2035.

Received on April 3, 1987

Note added in proof

These sequence data have been submitted to the EMBL/GenBank Data libraries under the accession number Y00070.

Another open reading frame, termed bel-3, of 167 amino acids runs from nucleotide 5250 to 5751. The protein sequence did not show any significant homology to the sequences available in the data banks. The bel-2 gene product contains one Arg-Gly-Asp sequence, a cell recognition signal crucial for interacting with its cell surface receptor, [Ruoshlat,E. and Pierschbacher,M.D. (1986) *Cell*, **44**, 517–518].