JOURNAL OF VIROLOGY, July 1987, p. 2182–2191
0022-538X/87/072182-10$02.00/0

Vol. 61, No. 7

Tissue-Specific Expression of Human Provirus ERV3 mRNA in Human Placenta: Two of the Three ERV3 mRNAs Contain Human Cellular Sequences

NOBUYUKI KATO,[1,†] SUSAN PFEIFER-OHLSSON,[2,‡] MIEKO KATO,[1] ERIK LARSSON,[3] JAN RYDNERT,[2] ROLF OHLSSON,[2‡] AND MAURICE COHEN[1,*]

BRI Basic Research Program, National Cancer Institute Frederick Cancer Research Facility, Frederick, Maryland 21701[1]; Department of Oncology, University of Umea, Umea, Sweden[2]; and Department of Pathology, Uppsala University, Uppsala, Sweden[3]

Received 23 December 1986/Accepted 24 March 1987

Three polyadenylated RNAs, 9, 7.3, and 3.5 kilobases long, of a human endogenous retrovirus, ERV3, are abundant in human placental chorion, representing about 0.03 to 0.05% of the total mRNA. We characterized the structure of these mRNAs by Northern blot and S1 nuclease mapping analyses. We found that all three RNAs were spliced mRNAs that lacked 5.9 kilobases of proviral sequence, including the *gag* gene and most of the *pol* gene. In contrast to the transcription pattern usual for other retroviruses, the transcription pattern of the ERV3 provirus did not include a genome-length mRNA. All three of the ERV3 mRNAs initiated transcription at the same point in the 5′ long terminal repeat (LTR) and contained identical splice junctions in the provirus. The 3.5-kilobase RNA was a typical subgenomic proviral mRNA, with its polyadenylation site in the 3′ LTR. The two larger ERV3 mRNAs, however, extended through the polyadenylation site in the 3′ LTR and were spliced at a second position approximately 370 nucleotides downstream from the 3′ LTR. This finding suggests that when the ERV3 retrovirus integrated at this genomic locus in an ancestor of humans, it integrated within or adjacent to a cellular gene.

Human DNA contains multiple copies of endogenous retroviral sequences related to mouse or primate retroviral sequences. Several human endogenous provirus families have been identified, and clones of each family have been isolated from human recombinant DNA libraries (4, 5, 8, 14–16, 18, 22, 23). Although three reports described RNA expression of a human type C-related provirus family (11, 27, 28), these reports could not identify the expressed gene locus because the provirus is present in 50 to 100 copies per haploid genome. We previously described the structure of a full-length human endogenous retrovirus, ERV3, which is present in a single copy per haploid genome and is located in chromosome 7 (21, 22). The nucleotide and encoded amino acid sequences of ERV3 were found to be significantly related to but distinct from those of Moloney murine leukemia virus (Mo-MuLV), baboon endogenous virus, and the human type C-related provirus family described previously (16). We recently reported the isolation of an ERV3 proviral cDNA clone from a human fetal liver library (6). In the present study, we demonstrate that ERV3 mRNAs of 9, 7.3, and 3.5 kilobases (kb) are expressed in the chorionic villi of human first-trimester and full-term placenta. Northern blot and S1 nuclease analyses revealed that these RNAs are all spliced, *env*-containing forms that have initiated at the same site in the 5′ long terminal repeat (LTR) and are missing identical sequences from the *gag* and *pol* genes. Although the 3.5-kb RNA is a typical subgenomic *env*-containing mRNA that terminates in the 3′ LTR with a poly(A) tail, the 9- and 7.3-kb ERV3 mRNAs have read through the 3′ LTR and are spliced again at a site downstream from the 3′ LTR.

* Corresponding author.
† Present address: National Cancer Center Research Institute, Tsukiji 5-chome, Chuo-ku, Tokyo, Japan.
‡ Present address: Centre for Biotechnology, Karolinska Institute, Huddinge, Sweden.

MATERIALS AND METHODS

Tissues. First-trimester human placentas were obtained from routine elective abortions by vacuum suction at the Department of Gynecology, Umea County Hospital, Umea, Sweden. Gestational age was determined by known date of conception or estimated from the date of the last menstrual period. In all cases gestational age was also evaluated by ultrasound investigation and from crown-rump length measurements. After immediate washing of the abortus, placental chorionic villi and whole embryos were isolated under a dissecting microscope. Full-term human placentas were obtained from Frederick Memorial Hospital, Frederick, Md. Tissues were immediately frozen in liquid nitrogen and stored at −70°C until needed.

RNA preparation. RNA was isolated from first-trimester placental tissues by homogenization in guanidinium isothiocyanate, CsCl gradient centrifugation, phenol-chloroform extraction, and ethanol precipitation (24). Total placental RNA from term placenta was prepared by a modification of the lithium chloride-urea procedure (1) as follows. Placental tissue was homogenized in a solution containing 10 volumes of 3 M lithium chloride–6 M urea, 50 mM Tris (pH 7.4), 5 mM EDTA, 0.1 M mercaptoethanol, and 0.1% (wt/vol) Sarkosyl by a Polytron homogenizer and then placed on ice overnight. After centrifugation at 8,000 rpm for 90 min at 4°C, the pellet was quickly suspended in a solution containing 10 mM Tris (pH 7.5), 1 mM EDTA, 0.5% (wt/vol) sodium dodecyl sulfate (SDS), and 200 μg of proteinase K per ml. The RNA was treated with phenol and precipitated with ethanol. After centrifugation the RNA was dissolved in 10 mM Tris (pH 7.5)–1 mM EDTA–0.5% (wt/vol) SDS, divided into portions, and stored in 70% ethanol at −20°C until used.

Northern blot hybridization. Total placental RNA (10 μg) was heated at 65°C for 10 min in 20 μl of 50% (vol/vol) formamide (Fluka)–2.2 M formaldehyde (Fisher Scientific)–20 mM MOPS (morpholinepropanesulfonic acid, pH 7.0)–5 mM sodium acetate–1 mM EDTA–0.2% (wt/vol) SDS and subjected to electrophoresis in a 0.8% agarose gel containing 2.2 M formaldehyde for 15 h at 40 V in buffer containing 20 mM MOPS (pH 7.0), 5 mM sodium acetate, and 1 mM EDTA. RNA was transferred overnight to a nitrocellulose membrane in 20× SSC (1× SSC is 0.15 M NaCl plus 15 mM sodium citrate) (32). The membrane was baked in a vacuum oven at 80°C for 5 h and soaked in a boiled solution of 20 mM Tris (pH 7.8) at room temperature for 10 min. The membrane was prehybridized for 5 to 7 h at 42°C in 50% (vol/vol) formamide–5× SSC–5× Denhardt solution (1× Denhardt solution is 0.02% [wt/vol] Ficoll, 0.02% [wt/vol] polyvinylpyrrolidone, 0.02% [wt/vol] bovine serum albumin)–80 μg of denatured yeast tRNA per ml–60 μg of denatured salmon sperm DNA per ml–0.1% (wt/vol) SDS. Nick-translated ^{32}P-labeled DNA fragment (specific activity 2×10^8 to 3×10^8 dpm/μg) was denatured by boiling and added to the hybridization solution, which was identical to the prehybridization solution except that the buffer contained 0.9 M NaCl, 50 mM sodium phosphate (pH 7.7), and 10% (wt/vol) dextran sulfate. The membrane-bound RNAs were annealed with probe (1.3×10^6 dpm/ml) for about 20 h at 42°C. After hybridization, the membrane was washed two times for 10 min each at room temperature in 2× SSC–0.1% SDS and two times for 1 h each at 68°C in 1× SSC–0.1% SDS or 0.1× SSC–0.1% SDS. Membranes were briefly dried and exposed to Kodak X-AR-5 film at −70°C with an intensifying screen. As molecular length markers, the RNA ladder (Bethesda Research Laboratories) was utilized. ERV3 DNA fragments used as hybridization probes were isolated from agarose gels by electrophoresis and binding to NA-45 membranes (Schleicher & Schuell) (35) and labeled with [α-^{32}P]dCTP (3,000 Ci/mmol; Amersham Inc.) by nick translation.

S1 mapping. The S1 mapping was performed as previously described (9), with some modifications. Total human placental RNA (25 μg) or yeast tRNA (25 μg) as a negative control was mixed well with ^{32}P-end-labeled DNA fragments (50,000 cpm [about 0.5 ng]) in 10 μl of hybridization solution (80% [vol/vol] formamide, 0.4 M NaCl, 40 mM HEPES [*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid] buffer [pH 6.4], 1 mM EDTA) and heated at 85°C for 10 min, immediately transferred to a suitable temperature estimated from the G+C content of the DNA probe, and incubated at that temperature overnight. This was followed by the addition of 0.3 ml of a 0°C solution containing S1 nuclease (Boehringer Mannheim) (200 to 400 U/ml) in 0.28 M NaCl–50 mM sodium acetate (pH 4.6)–4.5 mM zinc acetate–20 μg of denatured salmon sperm DNA per ml and incubated at 37°C for 30 min unless otherwise specified. The reactions were terminated by addition of 50 μl of a solution containing 4 M ammonium sulfate and 0.1 M EDTA. Nucleic acids were extracted and collected by ethanol precipitation. Samples were subjected to electrophoresis in an 8% polyacrylamide sequencing gel. Alternatively, protected fragments were denatured by glyoxal and subjected to electrophoresis in a 2% agarose gel containing 5% (wt/vol) glycerol.

RESULTS

Hybridization analysis of ERV3 transcripts from placenta. High levels of ERV3 *env*-containing RNAs were first identified in human first-trimester placental tissues (4.6 to 11 weeks gestational age) by dot blot hybridization analysis (Pfeifer-Ohlsson et al., unpublished results). Because ERV3 is a single-copy human provirus, ERV3 *env* probes detect a unique fragment in Southern blots washed at moderate stringency (22). In Northern blot analysis of RNAs isolated from the chorionic villi and embryos of human first-trimester abortuses, hybridization with an ERV3 *env* probe revealed a distinctive difference in transcription pattern (Fig. 1). Three ERV3 *env*-containing mRNAs of 9, 7.3, and 3.5 kb were detected in chorionic tissue; the 3.5-kb RNA was the most abundant. The 9-kb RNA was most abundant in embryos from these abortuses, with less of the 3.5-kb and almost none of the 7.3-kb species. RNA isolated from term placental chorion of twins (Fig. 1, lanes 22A and 22B) revealed a transcription pattern identical to that of first-trimester chorionic villi.

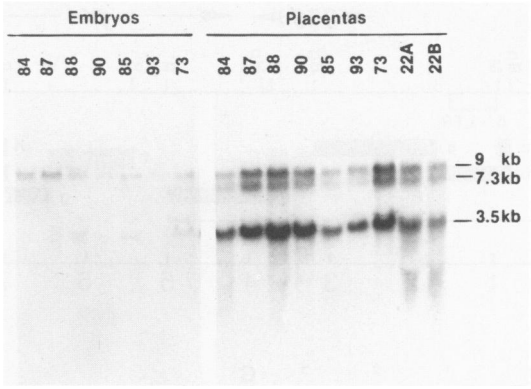


FIG. 1. Northern blot analysis of ERV3 mRNAs in human placentas. Total cellular RNA was isolated from placental chorionic villi (placentas) and embryos of individual first-trimester human placentas of 8 to 11 weeks gestational age (placentas 84, 87, 88, 90, 85, 93, and 73) or from the chorionic villi of twin full-term placentas (36 weeks plus 4 days) (placentas 22A and 22B). The hybridization probe was the ERV3 pBR322 subclone pHP 1.7, containing fragment j (Fig. 2).

To identify regions of the provirus expressed in ERV3 mRNAs, we used isolated fragments from the ERV3 lambda clone as probes in Northern blot hybridization to total RNA of full-term chorionic villi. The map locations of all probes used are shown in Fig. 2A. Probes b, c, i, j, k, l, m, n, o, and p resulted in a strong hybridization signal to all three ERV3 RNAs even after a high-stringency wash (0.1× SSC, 68°C) (Fig. 2B, lane 1). Probes d, e, f, g, and h failed to detect any of the ERV3 mRNAs (Fig. 2B, lane 2). These results suggested that the 9-, 7.3-, and 3.5-kb RNA species were all spliced RNAs. Because probes c and i detected the three ERV3 RNAs, whereas probes d and h were negative, we anticipated that all three RNAs would have the same splice junctions, a splice donor site in region c and a splice acceptor site in region i (region h excluded). Probes a, r, and s, which originated from outside the ERV3 provirus, did not detect the 9- or 7.3-kb (or 3.5-kb) RNAs, suggesting that the two long RNAs were not simply readthrough products. Probe q, however, detected the 9- and 7.3-kb mRNAs but not the 3.5-kb mRNA (Fig. 2B, lane 3). All three of the ERV3 RNAs were found to be concentrated in the poly(A)-containing [poly(A)$^+$] RNA fraction (Fig. 3), indicating that the 9- and 7.3-kb RNAs were not unprocessed precursors of the 3.5-kb species. In this and other experiments (unpublished), the 9-kb RNA appeared as a doublet. Because probe r did not

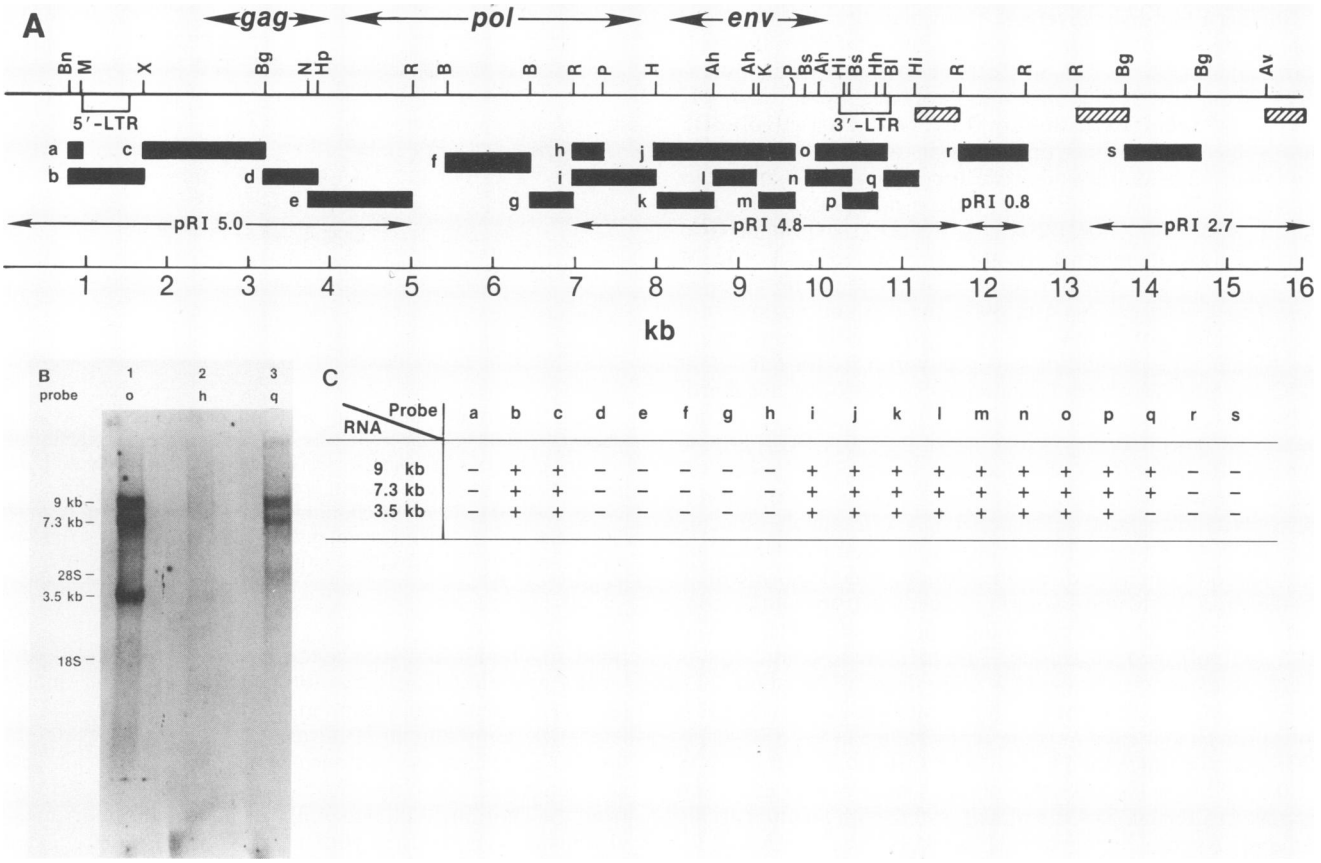
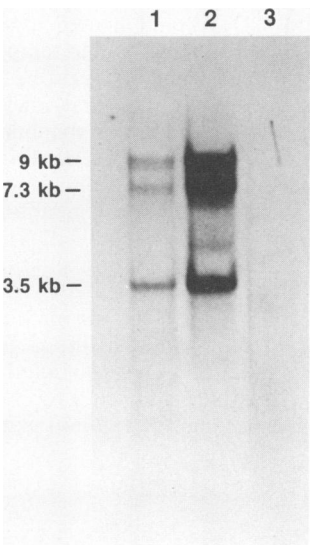


FIG. 2. Northern blot analysis of ERV3 mRNAs. (A) Map location of the probes in a human endogenous retroviral clone, ERV3. Fragments a (*Ban*I-*Mnl*I, 0.1 kb), b (*Ban*I-*Xma*I, 0.7 kb), c (*Xma*I-*Bgl*II, 1.4 kb), d (*Bgl*II-*Hpa*I, 0.6 kb), and e (*Nru*I-*Eco*RI, 1.3 kb) were isolated from pRI5.0 plasmid DNA (22). Fragments f (*Bam*HI-*Bam*HI, 1.0 kb) and g (*Bam*HI-*Eco*RI, 0.5 kb) were isolated from the ERV3 lambda clone. Fragments h (*Eco*RI-*Pst*I, 0.3 kb), i (*Eco*RI-*Hin*dIII, 1.0 kb), j (*Hin*dIII-*Pst*I, 1.7 kb), k (*Hin*dIII-*Aha*III, 0.8 kb), l (*Aha*III-*Aha*III, 0.6 kb), m (*Kpn*I-*Pst*I, 0.4 kb), n (*Aha*III-*Bal*I, 0.8 kb), o (*Bst*NI-*Bst*NI, 0.5 kb), p, (*Hin*fI-*Hha*I, 0.3 kb), and q (*Bal*I-*Hin*fI, 0.3 kb) were isolated from pRI4.8 plasmid DNA. Fragments r (*Eco*RI-*Eco*RI, 0.8 kb) and s (*Bgl*II-*Bgl*II, 0.8 kb) were obtained from pRI0.8 and pRI2.7 plasmid DNAs, respectively. Bars with cross-hatching represent regions that were strongly detected after hybridization with ^{32}P-labeled total human DNA, indicating that they contained highly repeated human sequences. Abbreviations: Ah, *Aha*III; Av, *Ava*I; B, *Bam*HI; Bg, *Bgl*II; Bl, *Bal*I; Bn, *Ban*I; Bs, *Bst*NI; H, *Hin*dIII; Hh, *Hha*I; Hi, *Hin*fI; Hp, *Hpa*I; K, *Kpn*I; M, *Mnl*I; N, *Nru*I; P, *Pst*I; R, *Eco*RI; X, *Xma*I. (B) Northern blot hybridization of human total placental RNA. Final washing condition was 0.1× SSC at 68°C (see Materials and Methods for additional details). Lanes: 1, fragment o as probe; 2, fragment h as probe; 3, fragment q as probe. (C) Summary of Northern blot analysis. +, Positive hybridization with the indicated probe.

RNA \ Probe	a	b	c	d	e	f	g	h	i	j	k	l	m	n	o	p	q	r	s
9 kb	−	+	+	−	−	−	−	−	+	+	+	+	+	+	+	+	+	−	−
7.3 kb	−	+	+	−	−	−	−	−	+	+	+	+	+	+	+	+	+	−	−
3.5 kb	−	+	+	−	−	−	−	−	+	+	+	+	+	+	+	+	−	−	−

hybridize to the ERV3 RNAs, these results suggested that the 9- and 7.3-kb RNAs either contained a second splice donor site or were polyadenylated in the 3′-flanking sequence near region q.

Initiation site of ERV3 mRNAs. We used S1 nuclease mapping to define the structure of ERV3 mRNAs which were estimated by the Northern blot analysis. To determine where the three ERV3 mRNAs initiated, we initially used a 1.6-kb *Eco*RI-*Xma*I fragment as an S1 nuclease hybridization probe (Fig. 4A, bottom). In this experiment we could not distinguish between protected probe and traces of remaining undigested probe (not shown). Therefore, we repeated the analysis, using as probe a 2.0-kb fragment composed of the 1.6-kb *Eco*RI-*Xma*I region plus 0.4 kb of plasmid sequence (Fig. 4A, bottom). The fragment was 5′-end labeled at the *Xma*I site. The predicted site of ERV3 transcription initiation is 171 base pairs (bp) upstream from the *Xma*I restriction site (21; unpublished results). If any of the ERV3 *env*-containing RNAs had initiated transcription upstream from the LTR, we would have expected to find protected fragments of 1.6 kb or less in addition to the



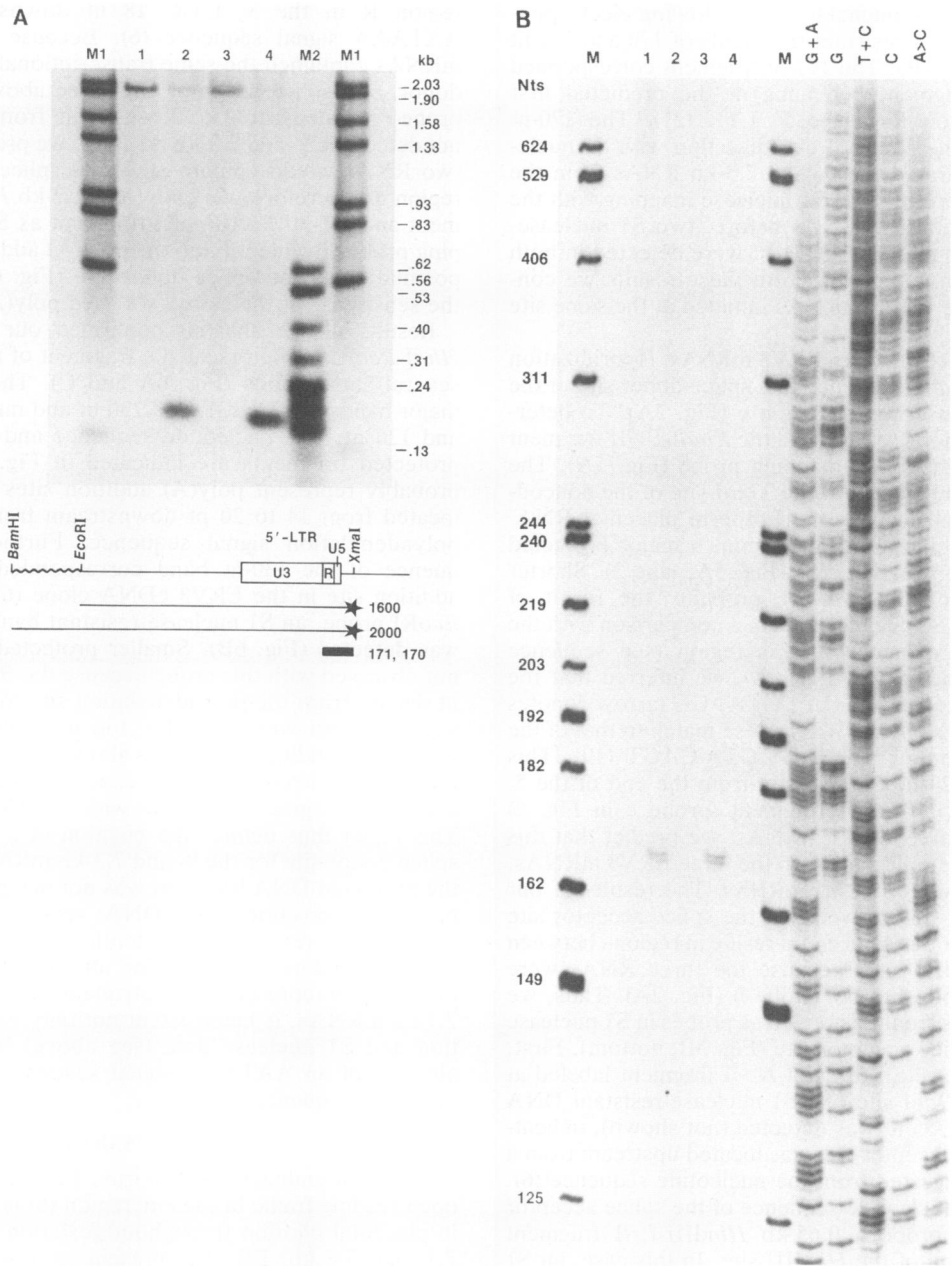


FIG. 4. S1 nuclease mapping analysis of ERV3 transcriptional initiation site. pRI5.0 plasmid DNA was digested with *Xma*I, dephosphorylated by bacterial alkaline phosphatase, and labeled at the 5′ end with [γ-^{32}P]ATP (3,000 Ci/mmol) and polynucleotide kinase. The DNA was recut with *Bam*HI, releasing a 2.0-kb fragment (*Xma*I-*Bam*HI) containing 0.4 kb of plasmid sequence. The labeled fragments were isolated by agarose gel electrophoresis. Thick bar at bottom corresponds to length of protected fragment in lanes 2 and 4. All S1 nuclease digests used 400 U of enzyme. (A) S1 nuclease-resistant fragments were detected by electrophoresis in a 2% agarose gel. Lanes: 1, yeast tRNA annealed at 52°C; 2, human full-term placental RNA annealed at 52°C; 3, yeast tRNA annealed at 55°C; 4, human full-term placental RNA annealed at 55°C. Lanes M1, ^{32}P 5′-end-labeled *Hin*dIII-*Eco*RI fragments of DNA as length markers; lane M2, ^{32}P 3′-end-labeled *Msp*I fragments of pBR322 DNA as length markers. (B) S1 nuclease-resistant fragments were detected by electrophoresis in an 8% polyacrylamide gel. Sequencing ladders were run for accurate fragment size measurement (17). Lanes: 1, yeast tRNA annealed at 52°C; 2, placental RNA annealed at 52°C; 3, yeast tRNA annealed at 55°C; 4, placental RNA annealed at 55°C; M, ^{32}P 3′-end-labeled *Msp*I fragments of pBR322 DNA used as length markers.

FIG. 3. Northern blot analysis after selection by oligo(dT)-cellulose. Poly(A)$^+$ RNA was isolated by one cycle of chromatography on oligo(dT)-cellulose (2). Lanes: 1, total RNA (10 μg); 2, poly(A)$^+$ RNA (5 μg); 3, poly(A)$^-$ RNA (10 μg). Autoradiographic exposure, 8 h.

171-nucleotide (nt) fragment. Using electrophoresis under conditions that clearly separated single-stranded DNA fragments of 2.0 and 1.6 kb, we found protected fragments of only 2.0 kb and approximately 170 nt (Fig. 4A). To obtain the precise nucleotide sequence at which transcription began, the protected DNA fragment was analyzed by electrophore-

sis in an acrylamide sequencing gel. Following electrophoresis, two S1 nuclease-resistant fragments of 170 and 171 nt were detected (Fig. 4B). The 171-nt fragment corresponded in length to a fragment beginning at the predicted first nucleotide of region R in the 5′ LTR (21). The 170-nt fragment may be the result of overdigestion with S1 nuclease. We then separated the 9- and 7.3-kb RNAs from the 3.5-kb RNA and repeated the S1 nuclease mapping with the fractionated placental RNAs. As before, two S1 nuclease-resistant fragments of 170 and 171 nt were detected in both RNA pools (data not shown). From these results we conclude that the three ERV3 mRNAs initiated at the same site in the 5′ LTR.

Splice donor site of the three ERV3 mRNAs. Hybridization analysis suggested the existence of a splice donor site of the ERV3 primary transcripts in region c (Fig. 2A). To determine the splice donor site, we used the *Xma*I-*Bgl*II fragment (1.4 kb) as an S1 nuclease mapping probe (Fig. 5A). The fragment was 3′-end labeled at the *Xma*I site of the noncoding strand and annealed to total full-term placental RNA. After S1 nuclease digestion for 30 min, a major protected fragment of 154 nt was detected (Fig. 5A, lane 2). Shorter bands in lane 4 of Fig. 5A are probably the result of overdigestion with S1 nuclease. From a comparison with the ERV3 nucleotide sequence in this region (see sequence ladder of noncoding strand, Fig. 5A), we inferred that the splice donor site sequence is CA ↓ GTAAGG (arrow denotes splice point). This sequence was a close match to that of the consensus splice donor site, AG ↓ GTA/GAGT (19). This site was located 224 nt downstream from the end of the 5′ LTR. Since the *Xma*I-*Bgl*II fragment (probe c in Fig. 2) hybridized to all three ERV3 mRNAs, we predict that this site was used as the splice donor in the three ERV3 mRNAs.

Splice acceptor site of ERV3 mRNAs. The results of our Northern blot analysis showed that the splice acceptor site of all three ERV3 transcripts could reside in region i between the *Pst*I and *Hin*dIII sites because the three RNAs were detected with probe i but not probe h (Fig. 2A). Thus, we used two fragments from this region as probes in S1 nuclease mapping of the splice acceptor site (Fig. 5B, bottom). First, we used as probe a 0.51-kb *Dde*I-*Nco*I fragment labeled at the 5′ end of the *Dde*I site. An S1 nuclease-resistant DNA fragment of about 355 nt was detected (not shown), indicating that the splice acceptor site was located upstream from a site previously suggested from the nucleotide sequence (6). To determine the nucleotide sequence of the splice acceptor site, we used as probe a 0.65-kb *Hin*dIII-*Pst*I fragment labeled at the 5′ end of the *Hin*dIII site. In this case, an S1 nuclease-resistant band of 190 nt was detected by polyacrylamide gel electrophoresis (Fig. 5B). This length exactly corresponded to that estimated from the *Dde*I-*Nco*I probe experiment. From the nucleotide sequence of this region (6), we predicted that the splice acceptor site of the three ERV3 mRNAs is CCACTGTTGTTAAAG ↓ T, which is located precisely 190 nt upstream from the labeled *Hin*dIII end. The splice acceptor site was located in the 3′ portion of the *pol* gene 469 nt upstream from the putative start of the *env* gene product (6).

Poly(A) addition site of 3.5-kb mRNA and second splice donor site of 9- and 7.3-kb mRNAs. As shown in Fig. 2C, probe p (*Hin*fI-*Hha*I) hybridized to the 9-, 7.3-, and 3.5-kb ERV3 mRNAs, but probe q (*Bal*I-*Hin*fI) hybridized to only the 9- and 7.3-kb mRNAs. Thus, we anticipated that the 3.5-kb mRNA would terminate in the 3′ LTR. Furthermore, an ERV3 cDNA clone obtained from a human second-trimester fetal liver library terminated at the estimated end of region R in the 3′ LTR, 18 nt downstream from the AATAAA signal sequence (6). Because all three ERV3 mRNAs contained the same transcriptional initiation, splice donor, and splice acceptor sites (see above), and because probe r (located 1 to 2 kb downstream from the 3′ LTR) did not detect the 9- and 7.3-kb mRNAs, we predicted that these two RNAs would contain a second splice donor site near region q. Therefore, we chose the 1.2-kb *Hha*I-*Eco*RI fragment and 1.1-kb *Bal*I-*Eco*RI fragment as S1 nuclease mapping probes for identifying the poly(A) addition site and the postulated second splice donor site (Fig. 6C). To increase the sensitivity of the assay, we used poly(A)$^+$ RNA.

Results of the analysis confirmed our prediction. The *Hha*I-*Eco*RI probe revealed a fragment of about 530 nt plus several small bands (Fig. 6A and C). The latter included major bands of 132, 131, and 130 nt and minor bands of 128 and 126 nt. The nucleotide sequence and termini of these protected fragments are indicated in Fig. 6D (21). These probably represent poly(A) addition sites since they were located from 14 to 20 nt downstream from the AATAAA polyadenylation signal sequence. Furthermore, the sequence of the 130-nt band corresponded to the poly(A) addition site in the ERV3 cDNA clone (6). With the *Bal*I-*Eco*RI probe, an S1 nuclease-resistant band of about 430 nt was detected (Fig. 6B). Smaller protected fragments were not observed with this probe because the *Bal*I site is only 31 nt distant from the polyadenylation site (6), and hybrids of such a length were probably too unstable to be detected under the conditions used. As shown in Fig. 6C, the end of the 430-nt protected fragment exactly coincided with that of the 530-nt fragment observed with the *Hha*I-*Eco*RI probe. This result thus defines the position of a probable second splice donor site for the 9- and 7.3-kb mRNAs, even though the protected DNA fragment was not mapped to the precise nucleotide position. The DNA sequence of the ERV3 lambda clone revealed two potential splice donor sites in this region (unpublished data). The alternative possibility, that this position represents the polyadenylation site of the 9- and 7.3-kb mRNAs, is inconsistent not only with the hybridization and S1 nuclease data (see above) but also with the absence of an AATAAA signal sequence in the adjacent, upstream sequence.

DISCUSSION

A human endogenous provirus, ERV3, which has a long open reading frame in the *env* region (6) is highly expressed in placental chorion throughout gestation as mRNAs of 9, 7.3, and 3.5 kb. ERV3 is present as a stable, single-copy provirus in the DNA of apparently all humans (over 20 individuals have been tested [22; N. Kato and M. Cohen, unpublished results]). This contrasts with other human proviruses, which are moderately or highly repetitive in human DNA (5, 8, 14–16, 23). Thus, we were able to investigate the detailed structure of the ERV3 mRNAs by using Northern blot hybridization and S1 nuclease mapping analysis.

In this report we characterized the structure of the ERV3 subgenomic 3.5-kb mRNA, including the initiation site, splice sites, and poly(A) addition site. We have determined that the 9- and 7.3-kb mRNAs are unusual in that they contain the same initiation and splice sites as the 3.5-kb mRNA but extend through the 3′ LTR and are spliced again about 370 nt downstream from the LTR. The probable structures of the three ERV3 mRNAs are summarized in Fig. 7.

A 5.9-kb *gag-pol* region was missing from all three ERV3 mRNAs, probably removed from the primary transcript by

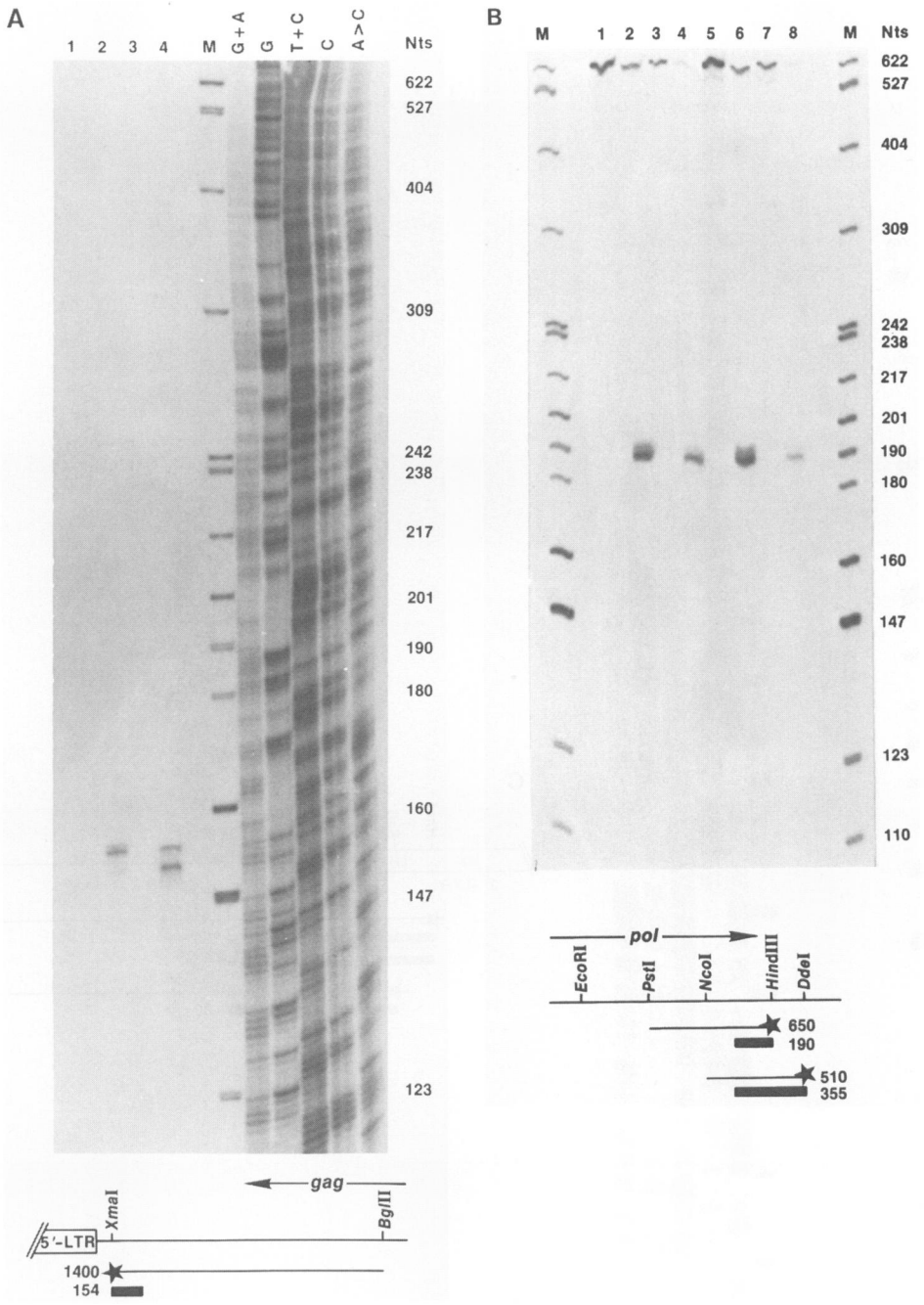


FIG. 5. S1 nuclease mapping analysis of ERV3 splice donor and acceptor sites. (A) Splice donor site. pRI5.0 plasmid DNA was digested with *Xma*I and labeled at the 3′ end with [α-^{32}P]dCTP (3,000 Ci/mmol) by using the Klenow fragment of DNA polymerase I and then recut with *Bgl*II. The 1.4-kb fragment probe (*Xma*I-*Bgl*II) was isolated by agarose gel electrophoresis. Hybridization was done at 63°C. Thick bar at bottom corresponds to length of major protected fragment in lane 2. S1 nuclease-resistant fragments were detected by 8% polyacrylamide gel electrophoresis. Sequencing ladder of the same *Xma*I end-labeled fragment was run to provide the DNA sequence around the splice junction. Lanes: 1, yeast tRNA, 400 U of S1 nuclease at 37°C, 30 min; 2, placental RNA, 400 U of S1 nuclease at 37°C, 30 min; 3, yeast tRNA, 400 U of S1 nuclease at 37°C, 90 min; 4, placental RNA, 400 U of S1 nuclease at 37°C, 90 min; M, ^{32}P-end-labeled *Msp*I fragments of pBR322 DNA as length markers. (B) Splice acceptor site. S1 nuclease mapping analysis of the ERV3 splice acceptor site. pSP65 plasmid DNA containing *Pst*I-*Sca*I fragment (−100 to 1109 [6]), subcloned from pRI4.8 plasmid, was digested with *Hin*dIII, dephosphorylated, and labeled at the 5′ end as described in legend to Fig. 4, and then recut with *Pst*I. The labeled *Hin*dIII-*Pst*I 0.65-kb fragment was isolated by agarose gel electrophoresis. Thick bar at bottom corresponds to length of major protected band. S1 nuclease-resistant fragments were detected by electrophoresis in a 8% polyacrylamide gel. Lanes: 1, yeast tRNA annealed at 54°C, 200 U of S1 nuclease; 2, placental RNA annealed at 54°C, 200 U of S1 nuclease; 3, yeast tRNA annealed at 54°C, 400 U of S1 nuclease; 4, placental RNA annealed at 54°C, 400 U of S1 nuclease; 5, yeast tRNA annealed at 59°C, 200 U of S1 nuclease; 6, placental RNA annealed at 59°C, 200 U of S1 nuclease; 7, yeast tRNA annealed at 59°C, 400 U of S1 nuclease; 8, placental RNA annealed at 59°C, 400 U of S1 nuclease; M, ^{32}P 5′-end-labeled *Msp*I fragments of pBR322 DNA used as length markers.

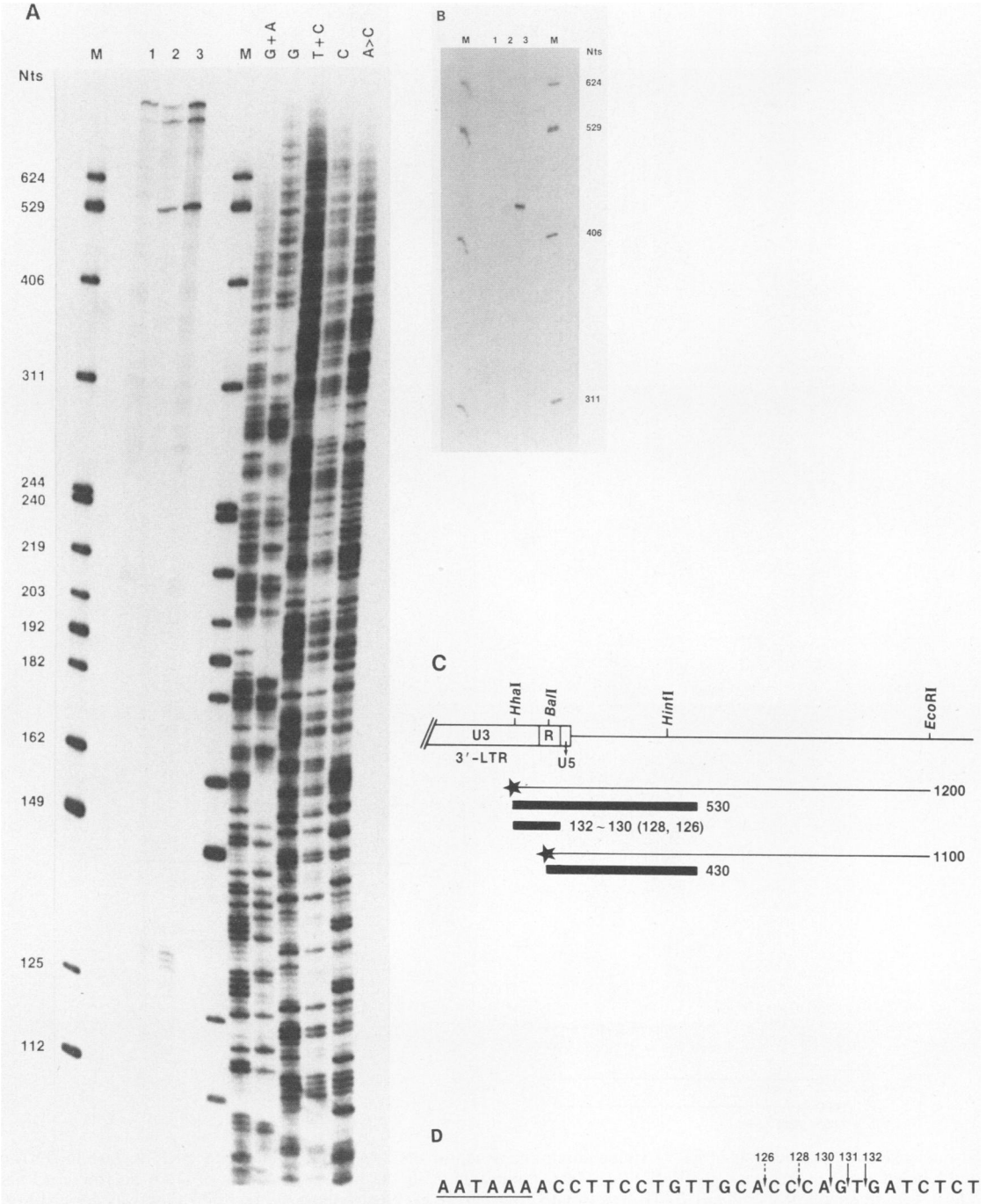


FIG. 6. S1 nuclease mapping analysis of the ERV3 polyadenylation site and second splice donor site. The *Hha*I-*Eco*RI 1.2-kb probe was obtained as follows. pRI4.8 plasmid DNA was digested with *Hha*I and labeled at the 3′ end with [α-^{32}P]dGTP (3,000 Ci/mmol) by using T4 DNA polymerase. After recleavage with *Eco*RI, the *Hha*I-*Eco*RI fragment was isolated by agarose gel electrophoresis. The *Bal*I-*Eco*RI 1.1-kb probe was obtained similarly. The pRI4.8 plasmid DNA was digested with *Bal*I and labeled at the 3′ end with the Klenow fragment of DNA polymerase I and [α-^{32}P]dCTP (3,000 Ci/mmol). After recleavage with *Eco*RI, the *Bal*I-*Eco*RI 1.1-kb fragment was isolated by agarose gel electrophoresis. (A) S1 nuclease-resistant fragments with the *Hha*I-*Eco*RI probe were detected by electrophoresis in an 8% polyacrylamide gel. Sequencing ladders were run for accurate fragment size measurement (17). Lanes: 1, yeast tRNA annealed at 55°C, 400 U of S1 nuclease; 2, total placental RNA annealed at 55°C, 400 U of S1 nuclease; 3, placental poly(A)$^+$ RNA (10 μg) annealed at 55°C, 400 U of S1 nuclease; M, ^{32}P 3′-end-labeled *Msp*I fragments of pBR322 DNA as size markers. (B) S1 nuclease-resistant fragment with the *Bal*I-*Eco*RI probe was detected by electrophoresis in an 8% polyacrylamide gel. Lanes and markers same as in panel A. (C) Map of the probes and S1 nuclease-resistant fragments. Thick bars represent lengths of protected fragments. (D) ERV3 sequence surrounding the poly(A) addition site(s). The arrows indicate the positions of polyadenylation.

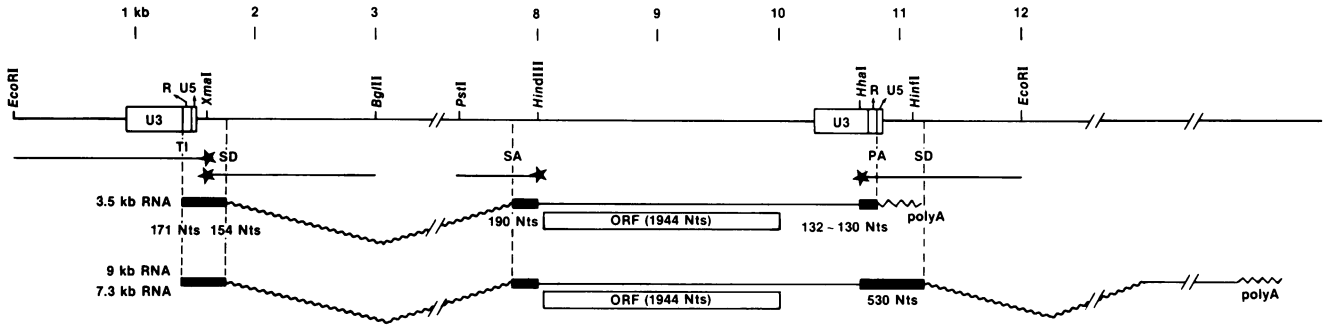


FIG. 7. Schematic structural presentation of the three ERV3 mRNAs. Each line with a star represents an end-labeled S1 nuclease mapping probe; black boxes indicate protected regions. ORF, Open reading frame; TI, transcription initiation site; SD, splice donor site; SA, splice acceptor site; PA, poly(A) addition site.

splicing (Fig. 5). Figure 8A shows the proviral sequences bounding the splice site of the ERV3 mRNAs. The intron sequence of the splice donor site closely matched that of the consensus sequence (19), although the exon ended with the rare dinucleotide CA. In addition, the intron sequence adjacent to the ERV3 splice acceptor site was unique. In contrast to the usual pyrimidine in the position 3 nt upstream from the splice acceptor site, ERV3 contained an adenosine. An adenosine at this position has been noted in only 3.0% of the published sequences (44 of 1,448 cases; data not shown) and in several different species. Interestingly, an adenosine in this position has also been found in the splice acceptor sequence of Mo-MuLV (31), human T-cell lymphotropic virus type I HTLV-I (29), and human c-*myc* intron 2 (26). The boundary sequence of human c-*myc* intron 2 is most homologous to the ERV3 splice acceptor site sequence (Fig. 8B). In the c-*myc* genes of chicken (34) and mouse (3), however, the c-*myc* splice acceptor sequence contains a cytosine at this position, and there is no evidence that changes in this nucleotide affect the specificity or rate of RNA processing.

The absence of an ERV3 genome-length mRNA is unusual. In contrast to the usual complete splicing of eucaryotic gene introns (10), processing of retroviral mRNAs is normally incomplete, resulting in cytoplasmic accumulation of a genome-length mRNA and a spliced subgenomic mRNA (33). ERV3 is atypical because the *gag-pol* intron was absent from all three mRNAs. This highly efficient ERV3 splice could be due either to the presence of a positive-acting cellular factor such as a small nuclear ribonucleoprotein or to tertiary folding of the primary transcript that would allow exceptional access to the splicing apparatus of the cell. Such possibilities were suggested for splicing of the P-element in *Drosophila melanogaster* (13).

An alternative explanation for the complete splicing of ERV3 transcripts is that a negative control element has been lost. The normal incomplete processing of retroviral transcripts may be due to the presence of a factor that specifi-

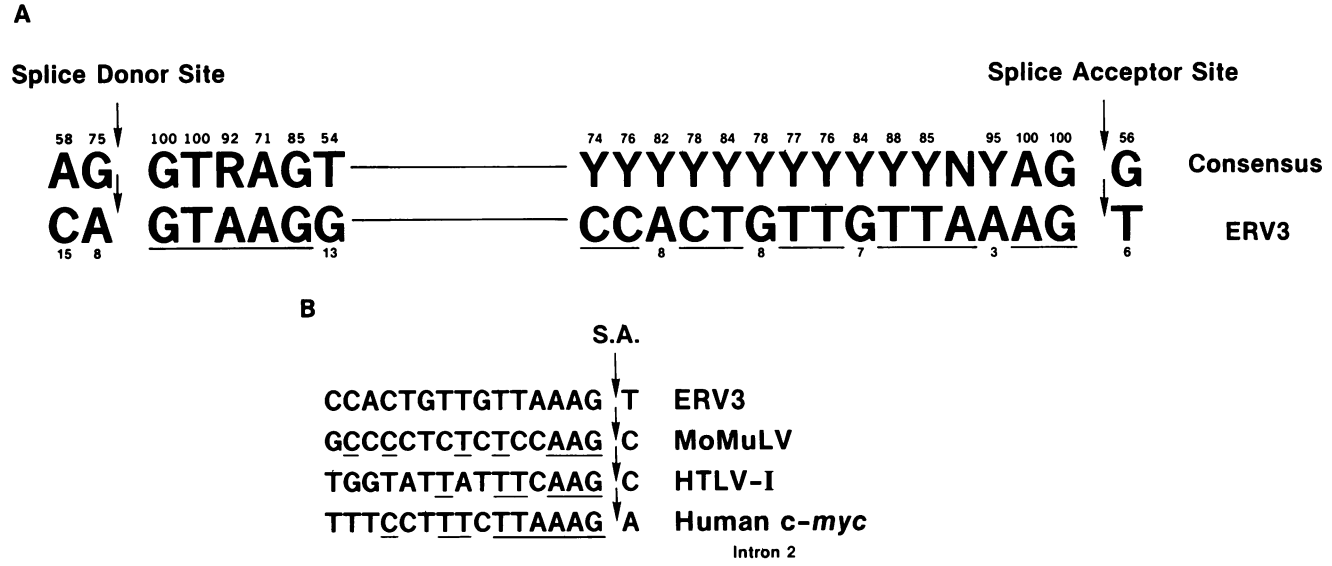


FIG. 8. Comparison of the ERV3 splice junction sequences with those of other genes. (A) Underlining: homologies between the ERV3 sequence and that of the consensus splice junction sequence of other genes (26, 29, 31). Numbers above and below sequences indicate the percentage of the known splice junctions containing that nucleotide at that particular position. (B) Underlining: splice acceptor site (S.A.) homology between ERV3 and other genes.

cally inhibits retroviral splicing. The absence of such a factor could explain why mammalian cells transformed with avian sarcoma virus expressed predominately a spliced *src* mRNA, whereas permissive avian cells expressed mainly the genome-length mRNA (25). From the experiments described in this report, it is not possible to decide whether splicing of the ERV3 *gag-pol* intron is controlled by a positive or a negative regulatory element, or whether the highly efficient splicing is simply a result of proviral mutations that have optimized splice site recognition.

We have precisely determined the poly(A) addition site of the 3.5-kb RNA species. In our S1 nuclease analysis (Fig. 6), three major S1 nuclease-resistant fragments were observed, the shortest of which (130 nt) mapped to the same location as the poly(A) addition site found in a cDNA clone isolated from a library made from RNA of a second-trimester human fetus (6). This site was 18 nt downstream from an AATAAA signal sequence. The two other major bands were equivalent to positions 19 and 20 nt downstream from the AATAAA sequence. Although multiple poly(A) addition sites may actually be present in the 3.5-kb ERV3 mRNA, the several S1 nuclease-resistant fragments may have originated from artifacts of the S1 nuclease digestion. Analysis of additional cDNA clones should resolve this question.

Poly(A) addition site analysis of the ERV3 mRNAs revealed a second splice donor site in the 9- and 7.3-kb mRNAs about 530 nt downstream from the *Hha*I-labeled end and 370 nt downstream from the 3′ LTR (Fig. 6A and C). A second S1 nuclease probe revealed a fragment that mapped to the identical location (Fig. 6B). This site cannot be the polyadenylation site of the 9- and 7.3-kb mRNAs because both species initiated at the same position as the 3.5-kb RNA and lacked the same 5.9-kb region, a result of splicing in the *gag-pol* genes. Thus, processing of the ERV3 primary transcript is different from that of the endogenous avian retrovirus, *ev*-1, that produces mRNAs polyadenylated 100 to 200 nt downstream of the normal poly(A) addition site in the 3′ flanking region (7). The ERV3 mRNAs, therefore, should contain an additional 5.5 or 3.8 kb of human sequence, respectively (the difference between 9 or 7.3 kb and 3.5 kb). If, in a human ancestor, the ERV3 provirus integrated in the coding sequence of an existing gene, the two large ERV3 mRNAs might be expected to contain several exons in the remaining sequence. Alternatively, a new splice donor site might have been fortuitously generated after integration.

Recently, novel processing of retroviral transcripts has been reported in two cases in which retroviruses have integrated within or adjacent to a cellular oncogene: avian leukosis virus (ALV)-induced erythroblastosis (12, 20) and Mo-MuLV-induced tumors (30). In the former case, integration of ALV into the c-*erbB* locus resulted in expression of different, spliced c-*erbB* mRNAs. One of these initiated in the 5′ LTR of the provirus and linked ALV *gag* sequences to the c-*erbB* gene, bypassing the 3′ LTR. Another mRNA which initiated in the 5′ LTR contained a double splice, from *gag* into *env* and from *env* into c-*erbB*, again bypassing the 3′ LTR (12, 20). In Mo-MuLV tumors, novel c-*myb* mRNAs initiating in the 5′ LTR of Mo-MuLV are spliced between the *gag* gene and exon 1 of the c-*myb* gene, removing *pol*, *env*, and the 3′ LTR (30). The 9- and 7.3-kb mRNAs of ERV3 are different from these examples because although the ERV3 mRNAs were linked to cellular sequences, the 3′ LTR was not bypassed via a splice.

This study demonstrates that ERV3 transcription is regulated at two levels. ERV3 expression is controlled at the level of transcriptional initiation or mRNA stability: ERV3 mRNA is significantly more abundant in the chorion than in the embryo (Fig. 1). Secondly, ERV3 expression is regulated at the level of RNA processing in two distinct ways: the three ERV3 mRNAs lacked sequences in the *gag-pol* region so that no genome-length mRNA was detected, and expression of the 9-kb and 7.3-kb mRNAs was regulated in the placenta because chorionic villi expressed about equal levels of these RNAs, whereas embryos expressed the 9-kb mRNA but almost none of the 7.3-kb form (Fig. 1). Although the ERV3 9-kb mRNA was detected in many normal human tissues, the 7.3-kb mRNA was observed primarily in the placental chorion (N. Kato et al., unpublished results). Because the two RNAs are apparently identical within their proviral portions, they must differ in their cellular domains as a consequence of alternative splicing or alternative poly(A) addition site selection. Analysis of cDNA and genomic clones downstream from the provirus should resolve the differences and make clear the significance of ERV3 mRNAs that contain human cellular sequences.

ACKNOWLEDGMENTS

We thank J. Cleveland, F. Propost, E. Brownell, and S. Hughes for critical reading of the manuscript.

Research was sponsored, in part, by the National Cancer Institute under contract NO1-CO-23909 with Bionetics Research, Inc. Research was also sponsored by the Japanese Overseas Cancer Fellow(ship) of the Foundation for Promotion of Cancer Research.

LITERATURE CITED

1. **Auffray, C., and F. Rougeon.** 1980. Purification of mouse immunoglobulin heavy-chain messenger RNAs from total myeloma tumor RNA. Eur. J. Biochem. **107**:303–314.
2. **Aviv, H., and P. Leder.** 1972. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA **69**:1408–1412.
3. **Bernard, O., S. Cory, S. Gerondakis, E. Webb, and J. M. Adams.** 1983. Sequence of the murine and human cellular *myc* oncogenes and two modes of *myc* transcription resulting from chromosome translocation in B lymphoid tumors. EMBO J. **2**:2375–2383.
4. **Bonner, T. I., C. O'Connell, and M. Cohen.** 1982. Cloned endogenous retroviral sequences from human DNA. Proc. Natl. Acad. Sci. USA **79**:4709–4713.
5. **Callahan, R., W. Drohan, S. Tronick, and J. Schlom.** 1982. Detection and cloning of human DNA sequences related to the mouse mammary tumor virus genome. Proc. Natl. Acad. Sci. USA **79**:5503–5507.
6. **Cohen, M., M. Powers, C. O'Connell, and N. Kato.** 1985. The nucleotide sequence of the *env* gene from the human provirus ERV3 and isolation and characterization of an ERV3-specific cDNA. Virology **147**:449–458.
7. **Conklin, K. F., J. M. Coffin, H. L. Robinson, M. Groudine, and R. Eisenman.** 1982. Role of methylation in the induced and spontaneous expression of the avian endogenous virus *ev*-1: DNA structure and gene products. Mol. Cell. Biol. **2**:638–652.
8. **Deen, K. C., and R. W. Sweet.** 1986. Murine mammary tumor virus *pol*-related sequences in human DNA: characterization and sequence comparison with the complete murine mammary tumor virus *pol* gene. J. Virol. **57**:422–432.
9. **Favaloro, J., R. Treisman, and R. Kamen.** 1980. Transcription maps of polyoma virus-specific RNA: analysis by two-dimensional nuclease S1 gel mapping. Methods Enzymol. **65**:718–749.
10. **Flint, J.** 1984. Processing of mRNA precursors in eukaryotic cells, p. 151–179. *In* D. Apirion (ed.), Processing of RNA. CRC Press, Boca Raton, Fla.
11. **Gattoni-Celli, S., K. Kirsch, S. Kalled, and K. J. Isselbacher.** 1986. Expression of type C-related endogenous retroviral sequences in human colon tumors and colon cancer cell lines. Proc. Natl. Acad. Sci. USA **83**:6127–6131.

12. **Goodwin, R. G., F. M. Rottman, T. Callaghan, H.-J. Kung, P. A. Maroney, and T. W. Nilsen.** 1986. c-*erbB* activation in avian leukosis virus-induced erythroblastosis: multiple epidermal growth factor receptor mRNAs are generated by alternative RNA processing. Mol. Cell. Biol. **6**:3128–3133.
13. **Laski, F. A., D. C. Rio, and G. M. Rubin.** 1986. Tissue specificity of Drosophila P element transposition is regulated at the level of mRNA splicing. Cell **44**:7–19.
14. **Maeda, N.** 1985. Nucleotide sequence of the haptoglobin and haptoglobin-related gene pair: the haptoglobin-related gene contains a retrovirus-like element. J. Biol. Chem. **260**:6698–6709.
15. **Mager, D. L., and P. S. Henthorn.** 1984. Identification of a retrovirus-like repetitive element in human DNA. Proc. Natl. Acad. Sci. USA **81**:7510–7514.
16. **Martin, M. A., T. Bryan, S. Rasheed, and A. S. Khan.** 1981. Identification and cloning of endogenous retroviral sequences present in human DNA. Proc. Natl. Acad. Sci. USA **78**:4892–4896.
17. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
18. **May, F. E. B., and B. R. Westley.** 1986. Structure of a human retroviral sequence related to mouse mammary tumor virus. J. Virol. **60**:743–749.
19. **Mount, S. M.** 1982. A catalogue of splice junction sequences. Nucleic Acids Res. **10**:459–472.
20. **Nilsen, T. W., P. A. Maroney, R. G. Goodwin, F. M. Rottman, L. B. Crittenden, M. A. Raines, and H.-J. Kung.** 1985. c-*erb*B activation in ALV-induced erythroblastosis: novel RNA processing and promoter insertion result in expression of an amino-truncated EGF receptor. Cell **41**:719–726.
21. **O'Connell, C. D., and M. Cohen.** 1984. The LTR sequences of a novel human endogenous retrovirus. Science **226**:1204–1206.
22. **O'Connell, C. D., S. J. O'Brien, W. G. Nash, and M. Cohen.** 1984. ERV3, a full-length human endogenous provirus: chromosomal localization and evolutionary relationships. Virology **138**:225–235.
23. **Ono, M.** 1986. Molecular cloning and long terminal repeat sequences of human endogenous retrovirus genes related to types A and B retrovirus genes. J. Virol. **58**:937–944.
24. **Pfeifer-Ohlsson, S., A. S. Goustin, J. Rydnert, T. Wahlstrom, L. Bjersing, D. Stehelin, and R. Ohlsson.** 1984. Spatial and temporal pattern of cellular *myc* oncogene expression in developing human placenta: implications for embryonic cell proliferation. Cell **38**:585–596.
25. **Quintrell, N., S. H. Hughes, H. E. Varmus, and J. M. Bishop.** 1980. Structure of viral DNA and RNA in mammalian cells infected with avian sarcoma virus. J. Mol. Biol. **143**:363–393.
26. **Rabbitts, T. H., P. H. Hamlyn, and R. Baer.** 1983. Altered nucleotide sequences of a translocated c-*myc* gene in Burkitt's lymphoma. Nature (London) **306**:760–765.
27. **Rabson, A. B., Y. Hamagishi, P. E. Steele, M. Tykocinski, and M. A. Martin.** 1985. Characterization of human endogenous retroviral envelope RNA transcripts. J. Virol. **56**:176–182.
28. **Rabson, A. B., P. E. Steele, C. F. Garon, and M. A. Martin.** 1983. mRNA transcripts related to full-length endogenous retroviral DNA in human cells. Nature (London) **306**:604–607.
29. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
30. **Shen-ong, G. L. C., H. C. Morse III, M. Potter, and J. F. Mushinski.** 1986. Two modes of c-*myb* activation in virus-induced mouse myeloid tumors. Mol. Cell. Biol. **6**:380–392.
31. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
32. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
33. **Varmus, H., and R. Swanstrom.** 1984. Replication of retroviruses, p. 369–512. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
34. **Watson, D. K., E. P. Reddy, P. H. Duesberg, and T. S. Papas.** 1983. Nucleotide sequence analysis of the chicken c-*myc* gene reveals homologous and unique coding regions by comparison with the transforming gene of avian myelocytomatosis virus MC29, *gag-myc*. Proc. Natl. Acad. Sci. USA **80**:2146–2150.
35. **Winberg, G., and M.-L. Hammarskjold.** 1980. Isolation of DNA from agarose gels using DEAE-paper. Application to restriction site mapping of adenovirus type 16 DNA. Nucleic Acids Res. **8**:253–264.