JOURNAL OF VIROLOGY, Nov. 1991, p. 5702–5709
0022-538X/91/115702-08$02.00/0

Vol. 65, No. 11

cDNA Sequence and Genomic Characterization of Intracisternal A-Particle-Related Retroviral Elements Containing an Envelope Gene

FRANK U. REUSS*† AND H. CHICA SCHALLER

Zentrum für Molekulare Biologie, University of Heidelberg, Im Neuenheimer Feld 282, D-6900 Heidelberg, Federal Republic of Germany

Received 25 March 1991/Accepted 23 July 1991

Intracisternal A-particle retrotransposons (IAPs) are retroviruslike elements that are defective in envelope protein synthesis and exist without an extracellular stage. We have isolated a novel class of cDNAs that are related to known IAP elements in the nucleotide and deduced protein sequence of *gag* and *pol* genes but also contain a previously unidentified reading frame between the *pol* gene and putative U3 region. Analysis of the deduced protein sequence reveals features of the putative protein that are characteristic of retroviral envelope proteins. The isolated cDNAs represent transcripts of multiple retroid elements in the mouse genome that were termed IAPE (intracisternal A-particle-related elements coding for envelope). IAPE *env* genes exist in approximately 200 copies per haploid genome as integral parts of the majority of these retroid elements. Four major IAPE subgroups could be distinguished after *Eco*RI digestion of genomic DNA.

Retroviruses and retrotransposons replicate via reverse transcription of an RNA intermediate. However, differences are found in their life cycles. Retroviruses are found as extracellular infectious particles and consist of RNA genomes protected by a protein core and a membranous envelope (38). Envelope proteins at the virion surface mediate entry of the virus into the cell by interaction with a cellular receptor, followed by fusion of viral and cellular membranes. After infection of a host cell, the retroviral RNA is reverse-transcribed into DNA and integrated into the host cell genome as a provirus. The structural organization of a prototype retroviral provirus is characterized by two long terminal repeats (LTRs) flanking the central coding region, which includes genes for the major internal structural protein (*gag*), RNA-dependent DNA polymerase (*pol*), and envelope glycoprotein (*env*).

Retrotransposons, in contrast, are repetitive genetic elements in the cellular genome that rely on reverse transcription for their amplification but do not encode envelope proteins and thus are thought to be unable to produce extracellular infectious particles (2). The genome of some retrotransposons is organized in a central coding region and two flanking LTRs, and thus they have structural and functional similarities with retroviral proviruses. Most of these retrotransposons are known to specify the synthesis of intracellular viruslike particles and can also be classified as retroviruses defective for particular steps in virus assembly. The copia elements of *Drosophila melanogaster*, for example, contain in the central coding region one open reading frame similar to the *gag* and *pol* genes of retroviruses (28) and give rise to intracellular viruslike particles that consist of *gag* gene product, reverse transcriptase, and full-length transcripts (35). The *Drosophila* elements gypsy, 17.6, and 297 contain, in addition to *gag* and *pol*, a third reading frame between the *pol* gene and the 3′ LTR (18, 25, 32). However, no similarity to retroviral *env* genes exists, and expression of this reading frame as well as particle formation has not been reported.

The intracisternal A-particle retrotransposons (IAPs) of rodents encode *gag* and *pol* proteins and direct the synthesis of viruslike particles that assemble at the cytoplasmic face of the endoplasmic reticulum membrane (see reference 21 for review). These particles consist of a *gag* precursor protein encapsidating genomic 7.2-kb RNA and acquire a membrane shell by budding into the cisternae of the endoplasmic reticulum. They are found in a variety of cell lines, tumors, early mouse embryos, and some tissues, especially the thymus. The budding process, however, seems to be incomplete in most IAPs, since the majority of them do not segregate from the membrane of the endoplasmic reticulum (31). Transport of IAPs along the cellular secretory pathway has not been observed. IAP elements exist as multiple copies per haploid genome and have been shown to alter expression of cellular genes after transposition (5, 11). Complete IAP elements were cloned from mouse and Syrian hamster genomes (26, 30). A considerable number of IAP elements carry large deletions of up to 3.7 kb in the *gag* and *pol* genes. In *Mus musculus*, a total of 1,000 copies per haploid genome include 700 full-length and 300 truncated elements. In all known IAP elements, the region between the *pol* gene and the 3′ LTR, which in retroviruses codes for *env* proteins, is drastically shortened by deletions and has multiple conserved stop codons present in all three reading frames. This region is considered to be the truncated and mutated vestige of a once-functional *env* gene. The loss of the *env* gene is thought to be the main reason for the absence of infectious extracellular IAP virions.

In this article, we report the sequence of a novel mouse cDNA family isolated from a neuroblastoma-glioma cell line. We investigated the structure and number of retroid elements in the mouse genome that are identical to the cDNAs isolated. The cDNAs are related to known IAP elements and transcripts but differ from these by containing additional coding capacity in the envelope region. The length and structural features of the predicted protein are characteristic

* Corresponding author.
† Present address: Department of Molecular Biology and Microbiology, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111.

of retroviral envelope proteins. This is the first report of a mammalian retrotransposon including an *env* gene, suggesting the existence of IAPE elements with close similarity to complete retroviral proviruses.

MATERIALS AND METHODS

Cell lines. NIH 3T3 fibroblasts and the mouse neuroblastoma × rat glioma × mouse neuroblastoma hybrid-hybrid cell line NH15-CA2 (13) were grown in Dulbecco's modified Eagle's medium containing 10% fetal calf serum. NH15-CA2 cells were obtained from B. Hamprecht and continuously propagated. Mouse pituitary-derived AtT20 cells (7) were cultured in RPMI 1640 medium (Boehringer Mannheim) supplemented with 10% fetal calf serum.

cDNA library and isolation of clones. cDNA libraries were prepared by G. Bilbe and A. Bach as follows. NH15-CA2 polyadenylated RNA was isolated from the cells by guanidinium thiocyanate extraction (6) followed by oligo(dT)-cellulose chromatography. cDNA was synthesized by the RNase H method (10) with either oligo(dT)$_{12-18}$ or random hexanucleotides (Pharmacia) as primers. The cDNAs were inserted into DNA expression vectors λgt11 and λgt10 (17) by the use of *Eco*RI linkers. The NH15-CA2 random-primed library in λgt11 was screened after induction with nitrocellulose filters soaked in 10 mM isopropyl-β-D-thiogalactopyranoside (Biomol) with neuropeptide head activator anti-idiotypic antiserum 64/10 (34) at a dilution of 1:100 as described before (17). Positive plaques were detected after incubating the filters with alkaline phosphatase-linked anti-rabbit immunoglobulin G (IgG) second antiserum (Promega) at a dilution of 1:7,000 for 2 h at room temperature by reaction with 330 μg of *p*-nitroblue tetrazolium chloride (Biomol) and 165 μg of 5-bromo-4-chloro-3-indolylphosphate (*p*-toluidinium salt; Biomol) per ml in alkaline phosphatase buffer (100 mM Tris hydrochloride [pH 9.5], 100 mM NaCl, 5 mM $MgCl_2$). Positive clones were further selected and tested by in situ hybridization of nick-translated cDNA fragments to fixed cells (1) for the presence in AtT20 cells and absence in NIH 3T3 cells of corresponding RNA. One clone with the correct expression pattern, designated IAPE-0, was used as nick-translated ^{32}P-labeled probe to isolate further clones by plaque hybridization (24). Clones IAPE-4 and IAPE-8 originated from the random-primed λgt11 library; IAPE-1, IAPE-10, and IAPE-28 were isolated from the oligo(dT)-primed library in λgt10.

DNA sequence analysis and generation of probes. Phage DNA was isolated from a liquid culture as described before (24), and the *Eco*RI cDNA fragments were subcloned into plasmid vector pGEM2 (Promega). DNA sequencing was performed on plasmid DNA by the dideoxy chain termination method (33) with modified T7 DNA polymerase (Sequenase; USB) according to the manufacturer's instructions. In the case of cDNA clone IAPE-1, unidirectional deletion constructs from both ends of *Eco*RI fragment IAPE-1 (1 to 3851) were created by the modified exonuclease III-S1 method (15) in plasmid pGEM2, and both DNA strands were sequenced completely. Several of these deletion clones were used for the generation of ^{32}P-labeled RNA probes specific for parts of IAPE-1. All probes were complementary to the sense strand of IAPE-1 and synthesized by RNA polymerases T7 or SP6 (Promega) following the procedure given by the manufacturer, after restriction digestion of the template plasmid. The names of the probes and the nucleotide positions of IAPE-1 covered are GAG (1 to 418), POL (1055 to 1270), ENV (2185 to 2441), LTR1 (3650 to 3851), and LTR2 (3846 to 4021). Control Probe pGEM2 PL represents the polylinker of the plasmid used (nucleotides 10 to 86).

Computer-assisted sequence analysis. Sequence analysis and comparisons were performed with the University of Wisconsin Genetics Computer Group sequence analysis software package (8). Programs FASTN and FASTP (23) were used to search nucleic acid and protein sequence data bases; MULTALIGN (19) was used to compare multiple sequences.

Isolation and analysis of genomic DNA. Genomic DNA was isolated from cell lines and liver tissue by the guanidinium-cesium chloride method described for isolation of RNA (24). After centrifugation onto a cesium chloride cushion, DNA was further purified in a continuous cesium chloride gradient with a mean density of 1.72 g/cm^3 (Beckman L8-M, VTI65.2, 62,000 rpm, 6 h, 20°C). When 2 μg of the preparation was applied to a gel, no RNA could be detected after ethidium bromide staining. Ten micrograms of each DNA was subjected to *Eco*RI or *Bam*HI digestion. ^{32}P-labeled *Eco*RI-*Hin*dIII-digested lambda DNA (as size markers) and defined amounts of IAPE-0 fragment (representing nucleotides 1929 to 2669 of IAPE-1) equivalent to 0.5, 1, 5, 20, 40, and 80 copies per haploid genome (5×10^{-18} mol of IAPE-0 per copy) were separated on the same agarose gel as the digested genomic DNA. The DNAs were transferred to a Hybond-N membrane (Amersham), hybridized for 16 h at 50°C in buffer containing 43% formamide, 5× SSC (1× SSC is 0.15 M sodium chloride plus 0.015 M sodium citrate), 5× Denhardt's solution (1 mg of Ficoll, 1 mg of polyvinylpyrrolidone, and 1 mg of bovine serum albumin per ml), 50 mM sodium phosphate (pH 7), 0.1% sodium dodecyl sulfate (SDS), and 0.2 mg of salmon sperm DNA per ml and washed three times for 30 min each in 1× SSC plus 0.1% SDS. After autoradiography, the probe was removed by incubating the membrane for 30 min, first in 0.4 N NaOH and then in a buffer consisting of 0.1× SSC, 0.1% SDS, and 0.2 M Tris hydrochloride (pH 7.5). Successful removal of the probe was tested by autoradiography.

Nucleotide sequence accession number. The nucleotide sequence data reported in this article have been given GenBank accession number M73818.

RESULTS

A novel IAP element-related family of cDNAs. In an attempt to clone a cDNA encoding an antigen previously defined on the surface of neural NH15-CA2 cells (34), we isolated and analyzed a cDNA clone from a λgt11 expression library of this rat-mouse neuroblastoma-glioma hybrid-hybrid cell line. Selection criteria included recognition by antiserum 64/10, presence of corresponding mRNA in AtT20 pituitary cells, and absence of corresponding mRNA in NIH 3T3 fibroblasts. The isolated cDNA clone, later designated IAPE-0, was sequenced and found to contain a continuous open reading frame throughout its sequence of 741 nucleotides (Fig. 1). It was therefore used as a probe in further screening procedures to find larger cDNA clones originating from the same transcript. Among 20 clones analyzed, the 4,021-bp cDNA IAPE-1 was the largest. Northern (RNA) blot analysis of NH15-CA2 mRNA indicated that IAPE-1 is an almost full-length copy of a cellular 4.2-kb mRNA (31a).

The DNA sequence of IAPE-1 (Fig. 1) and five other cDNAs was determined. Comparison of these cDNAs revealed that they are colinear (Fig. 2) and highly related but not identical. As a mean value, the sequences differed from each other at 2% of the nucleotides. This indicated the

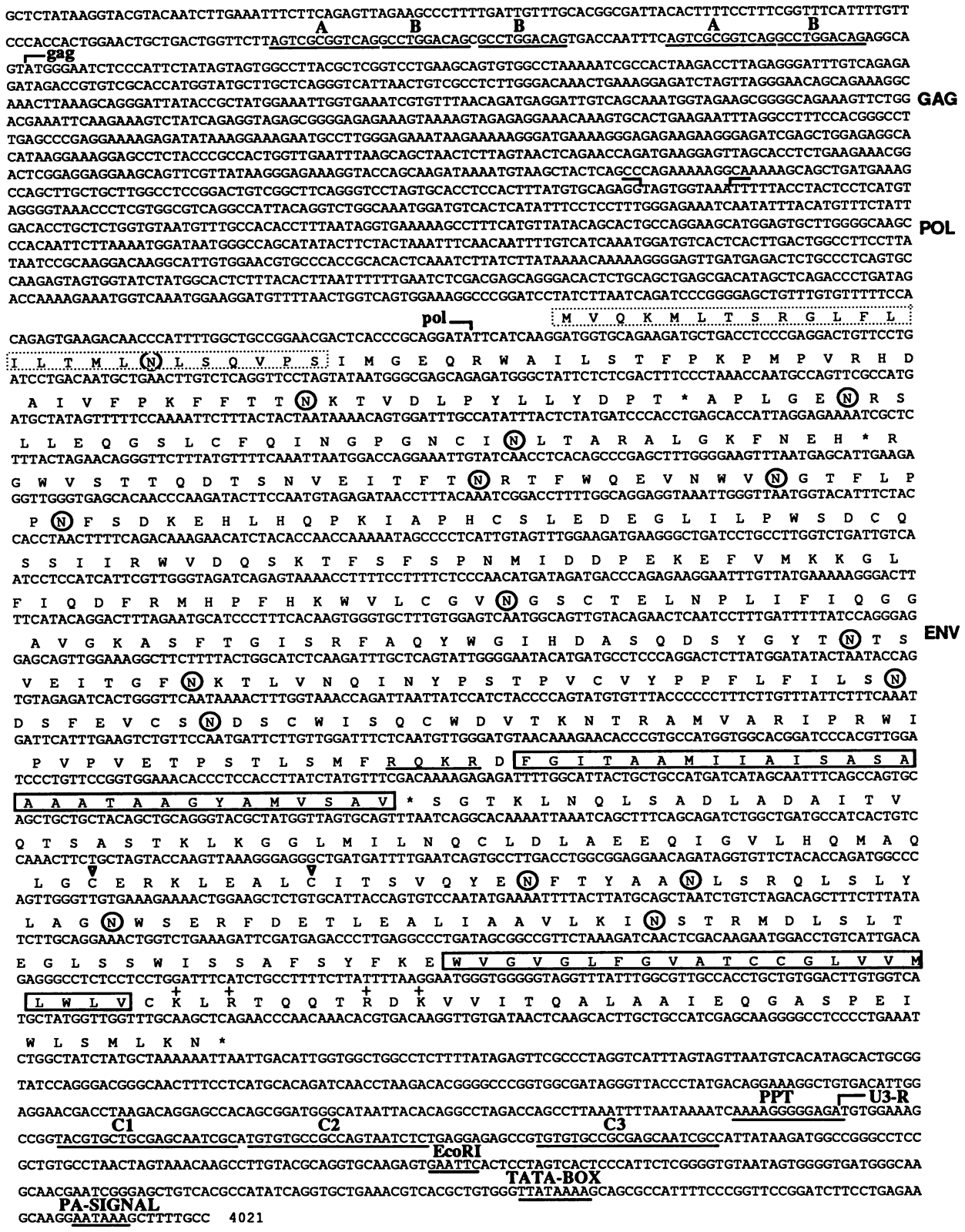

FIG. 1. Nucleotide sequence of IAPE-1 cDNA. The coding strand is shown. Reading frames identified by homology to IAP-coded proteins are marked and labeled *gag* or *pol*. Characteristic sequence elements are underlined: polypurine tract (PPT), TATA box, polyadenylation signal (PA signal), two 12-bp repeats (A), three 10-bp repeats (B), three imperfect 20-bp repeats (C1, C2, and C3), and the *Eco*RI restriction site. The suggested border of the U3 region is labeled. The sequence of cDNA IAPE-0 extends from nucleotides 1929 to 2669 and shows the nucleotide changes compared with IAPE-1 at positions 1995 (C), 1998 (G), 2167 (T), 2168 (C), 2175 (T), 2212 (C), 2387 (C), 2457 (G), and 2565 (C). The protein sequence deduced from the putative *env* gene of cDNA IAPE-1 is presented in single-letter code above the nucleotide sequence. Stop codons are represented by a star. The putative signal peptide (39) is boxed with a dotted line; other characteristic hydrophobic regions are boxed with a solid line. A potential protease cleavage site is underlined, potentially N-glycosylated Asn residues are marked by a circle, characteristic cysteine residues are marked by arrowheads, and basic residues adjacent to the proposed transmembrane region are marked by a plus sign.

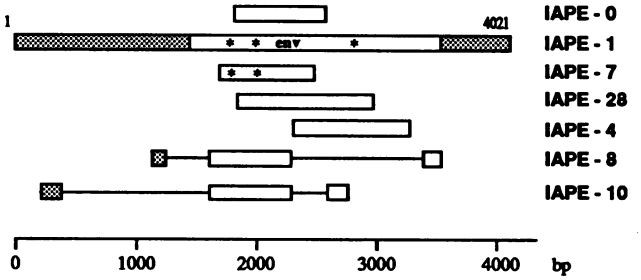

FIG. 2. Alignment of analyzed IAPE cDNAs. cDNAs are represented as boxes for regions with known sequence or as lines for regions of undetermined sequence. Putative *env* gene (open box) and other regions (stippled box) are indicated. Stars mark the position of stop codons in the *env* gene.

existence of a family of corresponding genes in the cellular genome. No sequence identical to that of cDNA IAPE-1 was found in the GenBank or EMBL data base, but strong nucleotide sequence similarities were obtained with several different IAP elements. Similarities with other retroid elements, retroviruses, and cDNAs were significantly lower. As shown in a dot matrix analysis of the DNA sequence homology between IAPE-1 and IAP element MIA14, the mouse IAP prototype sequence (Fig. 3) IAPE-1 has three distinct regions of similarity with MIA14. Nucleotides 475 to 970 are related to the *gag* gene, positions 981 to 1597 are related to the *pol* gene, and nucleotides 2900 to 3400 are related to the residual *env* region between the *pol* gene and the 3′ LTR of MIA14. The colinearity of these sequences is interrupted at two points. First, in IAPE-1, large parts of coding region equivalent to 3.8 kb of *gag* and *pol* genes in MIA14 are deleted. This type of truncation is found in approximately 30% of mouse IAP genomes (21). Second, a region with 1.25 kb of coding capacity, unknown in IAP element MIA14, is found in IAPE-1 located 3′ from the *pol*-homologous region. The sequence of the inserted DNA was found not to be related significantly to any sequence stored in public data bases.

The 5′- and 3′-terminal regions of IAPE-1 (nucleotides 1 to 475 and 3400 to 4021) are not related in sequence to MIA14. The region between nucleotides 1 and 200 upstream from the *gag* reading frame lacks open reading frames and is rich in repeat sequences (Fig. 1), as is the case in untranslated parts of IAP elements between the 5′ LTR and *gag* gene (26, 30).

Other characteristics allow the correlation of nucleotides 3680 to 4021 in IAPE-1 with the 3′ untranslated regions of IAP and retroviral RNAs. A polypurine tract is directly followed by the trinucleotide TGT, generally conserved at the 5′ end of IAP U3 regions. The presence of putative regulatory sequences usually concentrated in the LTRs of IAP elements and proviruses, such as a TATA box and polyadenylation signals, supports the classification of this region as a putative U3-R region of an LTR.

Proteins deduced from IAPE and IAP *gag* and *pol* genes are highly related. Comparison of the deduced amino acid sequences derived from homologous regions in IAPE-1 and IAPs indicates that the putative gene products are also related and suggests a common function (Fig. 4). The IAPE *gag* reading frame has 52% identity with the MIA14 IAP *gag* protein over a range of 160 amino acids and extends in the N-terminal direction for an additional 80 amino acids. The sequence is 61% similar to that of the full-length Syrian hamster IAP H18 (30) *gag* protein in these N-terminal amino

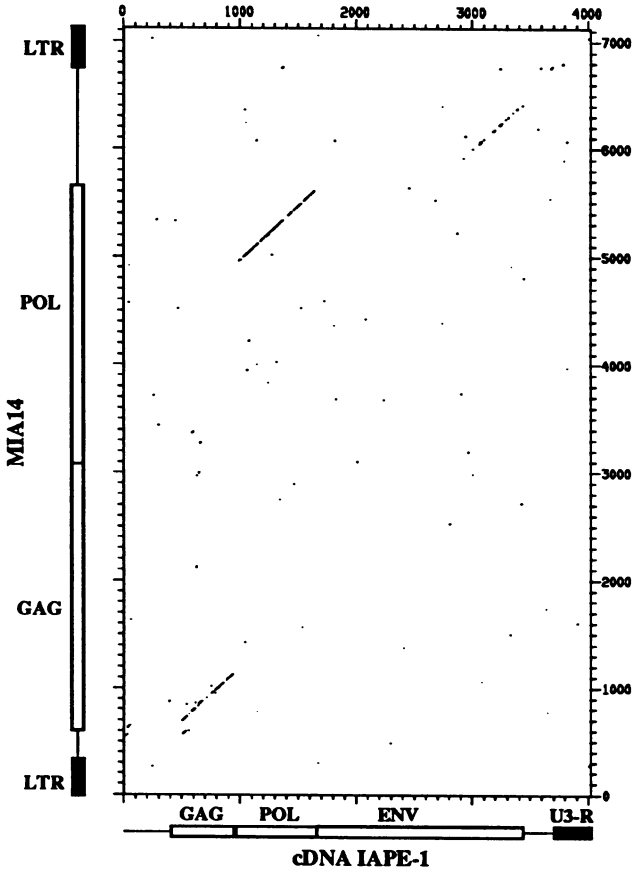

FIG. 3. Nucleotide comparison between IAPE-1 cDNA and mouse IAP element MIA14. Comparison was done by the program DOTPLOT (8). Each dot represents 75% or greater identity of a 20-nucleotide window in both sequences. Displayed at the vertical axis is the IAP element MIA14 (26). Coding regions (*gag*, *pol*) are represented by open rectangles; LTRs are shown as solid rectangles. The cDNA IAPE-1 is displayed at the horizontal axis. Regions with sequence similarity to MIA14 and putative *env* gene are marked by open rectangles. The nucleotide positions of the compared sequences are listed along the axes.

acids, but shows reduced similarity in the remainder of the sequence. The second amino acid in the putative IAPE-1 *gag* gene product is a glycine, known as a target for myristoylation in most retroviruses (12, 14).

The strongest similarities between IAPE-1 and IAP elements are found in the sequence deduced for the *pol* gene products. In cDNA IAPE-1, two adjacent reading frames separated by a single nucleotide deletion show similarities of 88 and 68% to the mouse IAP MIA14 and 86 and 79% to Syrian hamster IAP H18 *pol* protein.

IAPE cDNAs contain an unknown reading frame. All IAP elements are largely truncated between the *pol* gene and 3′ LTR, where the *env* gene is localized in retroviral proviruses. In cDNA IAPE-1, this region is extended by the presence of an additional 1.25 kb of sequence information. This unusual arrangement creates a novel 1.8-kb reading frame (nucleotides 1596 to 3422). The deduced amino acid sequence of this reading frame is shown in Fig. 1, starting with the first methionine. In IAPE-1, three stop codons are located at positions 72, 112, and 395. However, DNA sequence determination of several other IAPE cDNAs at

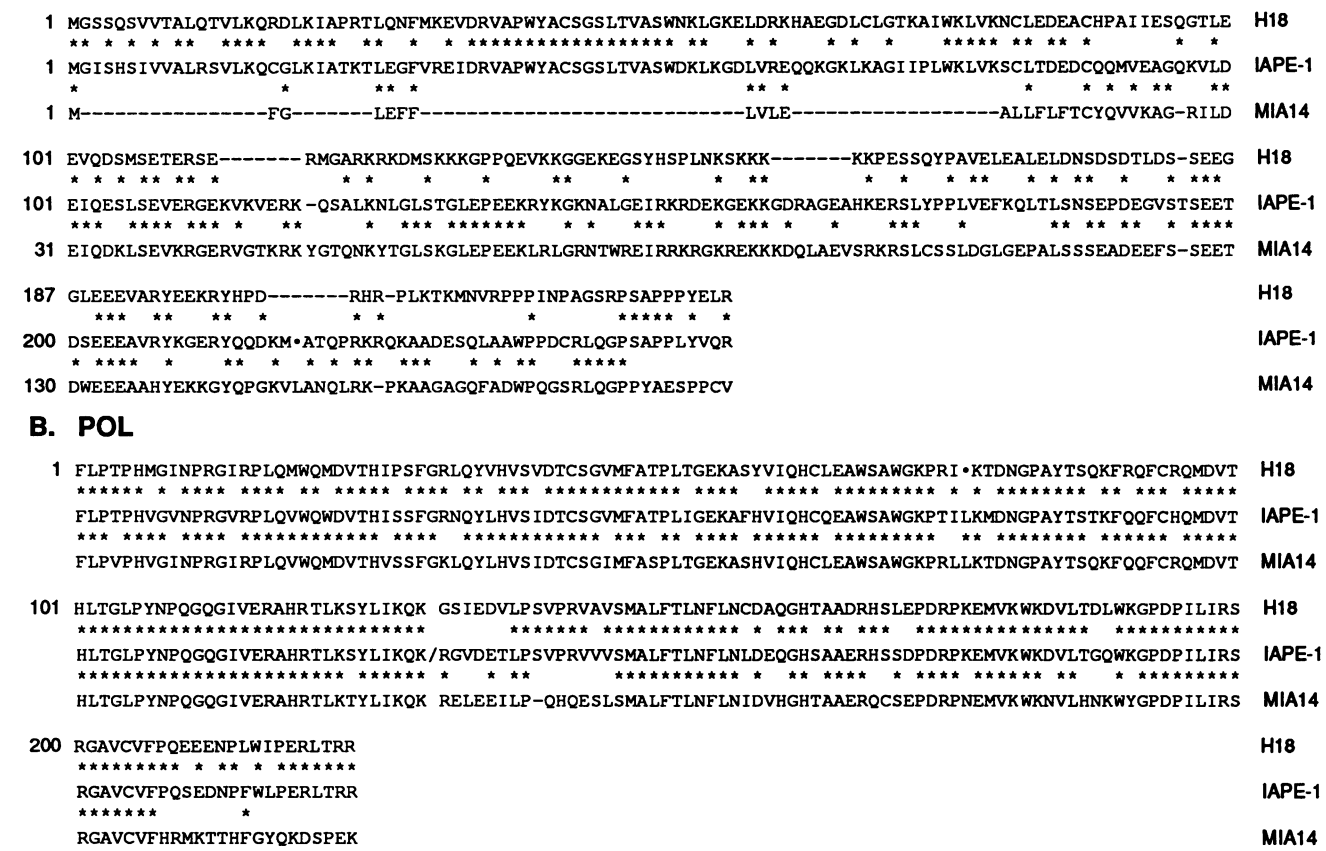

FIG. 4. Comparison between amino acid sequences deduced from the (A) *gag* and (B) *pol* genes of cDNA IAPE-1 and the genomic IAP elements mouse MIA14 (26) and Syrian hamster H18 (30). Introduced gaps are indicated by dashes, stop codons are shown by black dots, and a change of reading frame is shown by a slash. Amino acids identical in two of the sequences are marked by stars between the lines.

these positions revealed that these stop codons were absent in most IAPE clones sequenced (Fig. 2). Stop codons were replaced by arginine (positions 72 and 112) or glutamine (position 395) codons in five of seven IAPE cDNAs. These results led to the conclusion that this IAPE reading frame represents a potential gene that is interrupted by stop codons in some mRNAs. This gene could potentially direct the synthesis of a 587-amino-acid protein.

The unknown IAPE reading frame encodes an envelopelike protein. The protein sequence deduced from the novel IAPE gene (Fig. 1) was not found to exhibit significant similarity to any of the protein sequences deposited in public data bases. The localization of the gene between the *pol* gene and U3 region in the IAPE-1 cDNA suggests that it may code for a retroviral *env* protein. *env* proteins of different retroviruses exhibit little similarity in amino acid sequence but have common structural characteristics conditioned by transmembrane localization and common function (9, 16). Analysis of these characteristics (length, hydrophobicity profile, tryptic cleavage sites, potential N-glycosylation, and conserved cysteine residues) revealed that the postulated IAPE protein displays all the common structural characteristics of retroviral *env* proteins (9, 16). Length and hydrophobicity profiles coincide with the generally conserved patterns of *env* proteins, which are characterized by three hydrophobic peptides (signal peptide, fusion domain, and transmembrane region). Three hydrophobic regions are arranged in a particular order (Fig. 1). First, a hydrophobic N-terminal signal peptide (25 residues) is followed by a hydrophilic domain (340 residues). The end of this domain is marked by a cluster of basic residues, with the sequence Arg-Gln-Lys-Arg. This protease cleavage signal is adjacent to the second, residue 29, hydrophobic peptide. A second hydrophilic domain (135 residues) is bordered by a third hydrophobic peptide (22 residues). The C-terminal third hydrophilic domain is short (36 residues).

The length of the three hydrophilic *env* protein domains varies among different retroviruses. The hypothetical IAPE *env* protein, with hydrophilic regions of 340, 135, and 36 amino acids, resembles the B-type mouse mammary tumor virus (MMTV) *env* protein, comprising 358-, 137-, and 38-residue hydrophilic domains (27), although different lengths (436, 103, and 33 residues) are found in the *env* protein of the type C Moloney murine leukemia virus (36).

Retroviral *env* proteins are generally N-glycosylated at several asparagine residues within Asn-X-Ser/Thr consensus sequences. Sixteen such potential N-glycosylation sites are present in the sequence predicted for the novel IAPE protein and are localized exclusively in the region N-terminal to the proposed membrane-spanning domain. Another finding supporting an *env* function for the novel IAPE-encoded protein derives from the conservation of cysteine residues. In the described IAPE protein, a pair of cysteine residues (positions 450 and 458) are located 84 residues from the conserved

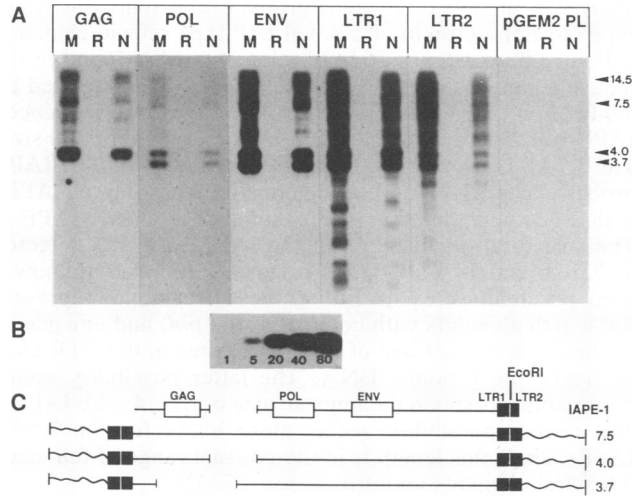

FIG. 5. Southern blot analysis of *Eco*RI-digested genomic DNA. (A) Genomic Southern blot with segment-specific IAPE probes. *Eco*RI-cleaved genomic DNA from NMRI mice (M), Wistar rats (R), and NH15-CA2 cells (N) was separated on a 1% agarose gel and transferred to a Hybond-N membrane. The membrane was successively hybridized to IAPE probes GAG, POL, ENV, LTR1, LTR2, and control probe pGEM2 PL and analyzed by autoradiography. pGEM2 PL contains all vector sequences present in the IAPE probes. The probe used is indicated above the lanes. Main fragments recognized by probe ENV1 are labeled by arrowheads, and the sizes of the fragments are given in kilobases. *Eco*RI-*Hin*dIII-digested λ DNA, radioactively labeled by filling the recessed ends with [α-^{32}P]dNTPs, was used as DNA size markers. (B) Quantitative control of genomic IAPE *env* copies. IAPE-0 cDNA (nucleotides 1929 to 2669 of IAPE-1) is recognized by probe ENV. Defined amounts of IAPE-0 cDNA equivalent to 1, 5, 20, 40, and 80 copies per haploid genome in 10 μg of genomic DNA (5×10^{-18} mol per copy) were run on the same gel as the genomic DNA, treated identically, and probed with ENV. The autoradiogram was scanned and compared with that in panel A. (C) Schematic representation of genomic IAPE elements as deduced from panel A. cDNA IAPE-1 is indicated, with the approximate positions of probes drawn as open boxes (GAG, POL, ENV) or black squares (LTR1, LTR2). Genomic IAPE elements are represented by a straight line; surrounding host DNA is shown by a waved line. The *Eco*RI restriction site used for digestion is marked by a vertical line. Numbers on the right indicate the size of the element in kilobases.

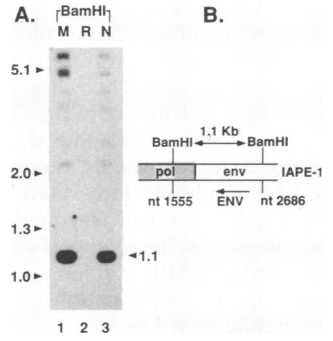

FIG. 6. Southern blot analysis of *Bam*HI-digested genomic DNA with IAPE probe ENV. (A) Genomic DNA from NMRI mice (M), Wistar rats (R), and cell line NH15-CA2 (N) digested with *Bam*HI was analyzed as described in the legend to Fig. 5, with ENV used as the ^{32}P-labeled probe. The sizes (in kilobases) and positions of ^{32}P-labeled *Eco*RI-*Hin*dIII-digested lambda DNA marker fragments are indicated at the left side, and the size of the main band is given at the right-hand side. (B) Schematic representation of the *Bam*HI sites near the *pol-env* gene junction in cDNA IAPE-1. The location of probe ENV is indicated by a singleheaded arrow. nt, nucleotide.

protease cleavage site. The *env* protein of MMTV reveals two identically spaced cysteines 85 residues from the cleavage site. In summary, these results suggest that the novel IAPE reading frame represents the IAPE *env* gene. Striking parallels between the deduced properties of the encoded protein and MMTV *env* proteins exist.

Multiple copies of the IAPE *env* gene are found in the mouse genome. All IAPE cDNAs were cloned from the rat-mouse fusion cell line NH15-CA2 and could be coded by either the rat or the mouse genome. In order to determine the species of origin for IAPE cDNAs, we performed a Southern blot analysis of rat and mouse genomic DNA, digested with restriction enzyme *Eco*RI. In the IAPE-1 cDNA, this enzyme recognized only one cleavage site that resides in the putative U3 region at position 3846 (Fig. 5C). For detection of IAPE *env* genes, we generated probe ENV (nucleotides 2185 to 2441 of IAPE-1).

Hybridization with probe ENV provided unequivocal evidence that the 1.25-kb region of cDNA IAPE-1, unknown from IAPs and representing part of the proposed *env* gene, was encoded by the mouse genome, not by the rat genome (Fig. 5A). *Eco*RI fragments were present in four major size classes, 14.5, 7.5, 4.0, and 3.7 kb. The use of different defined amounts of cloned cDNA IAPE-0 as a standard on the same gel (Fig. 5B) allowed us to calculate the approximate copy number of IAPE *env* genes per haploid genome after densitometric analysis of the autoradiograph. The fragments, in order of decreasing length, were present in approximately 25, 30, 50, and 50 copies. Including bands of minor intensity, a total of approximately 200 copies of the *env* gene per haploid mouse genome were detected. NH15-CA2 cells showed a fragment pattern very similar to that of the mouse genome; however, the intensity of individual bands was lower. This is probably related to the presence of rat DNA, which is devoid of IAPE elements, that exists in the NH15-CA2 genome and results in an apparent reduction in the number of IAPE copies.

The *env* gene is an integral part of the majority of genomic IAPE elements. To investigate whether all of the different putative *env* gene fragments are part of genomic IAPE elements, we hybridized IAPE-1 probes GAG, POL, LTR1, and LTR2 to the existing Southern blot (Fig. 5A). Hybridization of all these probes only to mouse DNA proved that all parts of the IAPE elements are encoded by the mouse genome. All major IAPE *env* gene-containing *Eco*RI fragments in the mouse genome were also recognized by other IAPE probes, indicating that the majority of detected *env* genes are part of IAPE elements. Fragments of 4.0 kb hybridized with all probes and corresponded in size to the cloned IAPE-1 cDNA. Thus, IAPE-1 cDNA could represent a transcript of these genomic elements. The 3.7-kb size class was recognized by all probes except GAG. High-molecular-weight fragments of approximately 14.5 and 7.5 kb were seen with all probes. Southern blot hybridization of *Bam*HI-digested genomic DNA to the probe ENV (Fig. 6) resulted in the detection of one major 1.1-kb band. This size is exactly that predicted from the IAPE-1 cDNA sequence and thus confirms the physical linkage of the *env* gene to the IAPE *pol* gene in the majority of IAPE copies.

When LTR probes were used on *Eco*RI-digested genomic DNA (Fig. 5A), a number of fragments detected were not

recognized by the GAG, POL, or ENV probes. These fragments probably represented the terminal parts of IAPE LTRs that stay with the neighboring host DNA sequences after *Eco*RI digestion (Fig. 5C). Their size distribution gave some clues concerning the structure of the DNA surrounding the IAPE elements. In both mouse and NH15-CA2 genomes, these DNAs did not exhibit a totally random size distribution, as expected for randomly integrated IAPE elements, but were concentrated in a set of fragments with different, distinct sizes and intensities. This suggests either a tandemly clustered arrangement of IAPE elements at several sites in the mouse genome or an amplification of IAPE DNA together with surrounding host DNA.

DISCUSSION

IAPs are retrotransposons which have a strong impact on the genome of many rodent species. A large family of endogenous elements encodes the particle components and is a frequent cause for activation or inactivation of cellular genes (5, 11). All previously cloned IAP elements or transcripts lack the *env* gene encoded by retroviruses (26). Consistent with this lack of *env*-coding information, an IAP-specific *env* protein, extracellular IAP virions, and horizontal transmission of IAPs via infection of other individuals are not known. We have characterized a novel class of mouse cDNAs and genomic elements related to IAP elements that contain an envelope gene. We designate these IAPE for intracisternal A-particle-related elements coding for envelope.

IAPEs are related to known IAP elements (21) in their genome organization, nucleotide sequence, putative *gag* and *pol* proteins encoded, the presence of multiple copies in the mouse genome, and the existence of truncated and mutationally inactivated coding sequences. Additional support for the relationship between IAPs and IAPEs comes from the identification of a 20-bp region complementary to mouse $tRNA^{Phe}$ in IAPEs (31a). This tRNA species is regarded as the primer involved in the synthesis of negative-strand cDNA during replication in mouse and hamster IAPs and has not been shown to be used by any other retrovirus or retrotransposon (29). However, the low degree of homology between the *gag* and *pol* proteins and LTR sequences of IAPs and IAPEs suggests that these represent distinct families of retrotransposons.

The most important characteristic of IAPEs is the presence of additional sequence information between the *pol* gene and U3 region, not found in previously characterized IAP elements. Together with sequences also common to IAPs, this novel DNA sequence encodes a product which structurally resembles retroviral *env* proteins. The *env* gene product of MMTV displays a very similar structural organization, although the amino acid sequences are not related (27). The presence of such a putative *env* gene in IAPEs is unique among the known mammalian retrotransposons (2). Only in *D. melanogaster* have retrotransposons been found that carry a third reading frame at an equivalent position. The deduced product lacks obvious sequence similarity to retroviral *env* proteins (18, 25, 32). The putative IAPE *env* gene is present in approximately 200 copies per haploid mouse genome and is linked to IAPE sequences. Thus, the *env* gene is an integral part of the majority of IAPEs. However, because of the large number of IAPEs and possible restriction fragment polymorphisms, we cannot exclude the presence of single *env* genes unlinked to IAPEs. It seems likely that the IAPE *env* gene product represents the prototype envelope protein of a novel IAP-type retroid element.

Four different types of IAPE elements were found in the mouse genome. Genomic IAPE elements, characterized by 4.0-kb *Eco*RI fragments, are equivalent to the sequenced IAPE-1 cDNA, as judged by hybridization pattern and size. The 3.7-kb IAPE fragments can be detected by all the IAPE probes except GAG and hence constitute a group of IAPEs with a *gag* gene deletion larger than that in cDNA IAPE-1. The structural organization of the IAPE elements detected as 7.5-kb and 14.5-kb *Eco*RI fragments remains unknown. Two possibilities exist. Either these fragments represent full-length elements with complete *gag*, *pol*, and *env* genes, or they lack at least one of the *Eco*RI sites in the LTR used to cleave the genomic DNA. The latter possibility seems reasonable to explain the unusual size of the 14.5-kb IAPEs. The former possibility seems more likely for the 7.5-kb IAPEs, since this length is in the unusual range of full-sized retroviral elements and proviruses (25, 37).

One of the possible transcripts of IAPE elements is represented by cDNA IAPE-1. Since the *gag* and *pol* genes are truncated, synthesis of functional products from both genes can be excluded. The existence of similarly truncated IAP elements and transcripts is known (21). The use of this transcript for *env* protein expression is unlikely for two reasons. First, retroviral *env* proteins are generally translated from subgenomic mRNAs from which *gag* and *pol* gene-encoded residues have been specifically removed by a splicing event. In IAPE-1, however, 1.45 kb of the *gag* and *pol* genes are still present. Second, the *env* reading frame in this cDNA is interrupted by three stop codons and is probably nonfunctional. From the sequence of other IAPE cDNAs, we know that all three stop codons in the IAPE-1 *env* gene are not conserved in other IAPEs but code for specific amino acids at the same position. This is in contrast to IAP elements, in which the majority of stop codons in the residual *env* region are conserved between different elements (26). Although we cannot at present report the isolation of a functional *env* cDNA, detection of intracellular IAPE *env* proteins with a sequence-specific antiserum (31a) makes the existence of intact *env* genes highly probable. The complete translation product of this gene, starting with the first methionine codon, has a calculated molecular mass of 65.3 kDa. After subtraction of the first 25 amino acids that may be the putative signal peptide (39), a size of 62.0 kDa is expected for the unglycosylated protein. Such an IAPE *env* protein could enable IAPEs to leave the host cell as a retrovirus that could be an as yet unknown infectious agent. Expression of IAPE *env* proteins in the endoplasmic reticulum may be sufficient to rescue the accumulating IAPs and create an infectious virus.

Two infectious murine retroviruses must be considered possible relatives of IAPEs. M432 virus was isolated from *Mus cervicolor* cells by cocultivation with cell line NIH 3T3 (3). A genetic relationship between the IAP and M432 genomes was indicated by hybridization experiments (22) and mapped to the *pol* gene by heteroduplex analysis (4). The antigenic relationships that exist between major structural proteins supported this view (20). A partial sequence of an M432 provirus, mainly encompassing the 3′ LTR, was kindly provided by K. K. Lueders. A comparison of this sequence with those of the IAPEs failed to detect any significant homologies. Thus, M432 and IAPEs are not identical but may be related in coding regions. The second virus with similarities to IAPEs is MMTV. The above-mentioned homology in structural organization of *env* pro-

teins implies similarities in assembly and structure of a hypothetical IAPE virus particle with MMTV virions. Analysis of IAPE *env* protein expression and the search for extracellular virions are currently in process.

ACKNOWLEDGMENTS

We thank U. Klingmüller for encouragement, G. Bilbe and A. Bach for preparation of cDNA libraries, G. Bilbe for isolation of IAPE-0, S. Hoffmeister for teaching in situ hybridization, and H. Schaller and R. M. Flügel for helpful discussions. We also thank K. K. Lueders for kind and informative correspondence, W. Reiser and A. Helfer for synthesis of oligonucleotides, and H. Schaller, M. Fischer, and L. Runkel for critical reading of the manuscript.

This work was supported by DFG (SFB 317) and by Fond der chemischen Industrie.

REFERENCES

1. **Berger, C. N.** 1986. In situ hybridisation of immunoglobulin-specific RNA in single cells of the B lymphocyte lineage with radiolabelled DNA probes. EMBO J. **5**:85–93.
2. **Boeke, J. D., and V. D. Corces.** 1989. Transcription and reverse transcription of retrotransposons. Annu. Rev. Microbiol. **42**: 403–434.
3. **Callahan, R., R. E. Benveniste, C. J. Sherr, G. Schidlowsky, and G. J. Todaro.** 1976. A new class of genetically transmitted retravirus isolated from *Mus cervicolor*. Proc. Natl. Acad. Sci. USA **73**:3579–3583.
4. **Callahan, R., E. L. Kuff, K. K. Lueders, and E. Birkenmeier.** 1981. Genetic relationship between the *Mus cervicolor* M432 retrovirus and the *Mus musculus* intracisternal type A particle. J. Virol. **40**:901–911.
5. **Canaani, E., O. Dreazen, A. Klar, G. Rechavi, D. Ram, J. B. Cohen, and D. Givol.** 1983. Activation of the c-*mos* oncogene in a mouse plasmacytoma by insertion of an endogenous intracisternal A-particle genome. Proc. Natl. Acad. Sci. USA **80**:7118–7122.
6. **Chomczynski, P., and N. Sacci.** 1987. Single step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. **162**:156–159.
7. **Cohen, A. I., J. Furth, and R. F. Buffet.** 1975. Histologic and physiologic characteristics of hormone secretory transplantable adrenal tumours in mice and rats. Am. J. Pathol. **33**:631–651.
8. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
9. **Gallaher, W. R., J. M. Ball, R. F. Garry, M. C. Griffin, and R. C. Montelaro.** 1989. A general model for the transmembrane proteins of HIV and other retroviruses. AIDS Res. Hum. Retroviruses **5**:431–440.
10. **Gubler, U., and B. J. Hoffman.** 1983. A simple and very efficient method for generating cDNA libraries. Gene **25**:263–269.
11. **Hawley, R. G., M. J. Shulman, and N. Hozumi.** 1984. Transposition of two different intracisternal A particle elements into an immunoglobulin kappa-chain gene. Mol. Cell. Biol. **4**:2565–2572.
12. **Henderson, L. E., C. Krutzsch, and S. Oroszlan.** 1983. Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational protein modification. Proc. Natl. Acad. Sci. USA **80**:339–343.
13. **Heumann, R., M. Öcalan, V. Kachel, and B. Hamprecht.** 1979. Clonal hybrid cell lines expressing cholinergic and adrenergic properties. Proc. Natl. Acad. Sci. USA **76**:4674–4677.
14. **Hizi, A., L. E. Henderson, T. D. Copeland, R. C. Sowder, H. C. Krutzsch, and S. Oroszlan.** 1989. Analysis of *gag* proteins from mouse mammary tumor virus. J. Virol. **63**:2543–2549.
15. **Hoheisel, J., and F. M. Pohl.** 1986. Simplified preparation of unidirectional deletion clones. Nucleic Acids Res. **14**:3605.
16. **Hunter, E., and R. Swanstrom.** 1990. Retrovirus envelope proteins. Curr. Top. Microbiol. Immunol. **157**:187–253.
17. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Constructing and screening of libraries in λgt10 and λgt11, p. 49–78. *In* D. M. Glover (ed.), DNA cloning, vol. 1. IRL Press, Oxford, England.
18. **Inouye, S., S. Yuki, and K. Saigo.** 1986. Complete nucleotide sequence and genome organisation of a Drosophila transposable element, 297. Eur. J. Biochem. **154**:417–425.
19. **Krueger, M., and G. Osterburg.** 1983. On the alignment of two or more molecular sequences. Comp. Prog. Biomed. **16**:61–70.
20. **Kuff, E. L., R. Callahan, and R. S. Howk.** 1980. Immunological relationship between the structural proteins of intracisternal A-particles of *Mus musculus* and the M432 retrovirus of *Mus cervicolor*. J. Virol. **33**:1211–1214.
21. **Kuff, E. L., and K. K. Lueders.** 1988. The intracisternal A-particle gene family: structure and functional aspects. Adv. Cancer Res. **51**:183–276.
22. **Kuff, E. L., K. K. Lueders, and E. M. Scolnick.** 1978. Nucleotide sequence relationship between intracisternal type A particles of *Mus musculus* and endogenous retrovirus (M432) of *Mus cervicolor*. J. Virol. **28**:66–74.
23. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
24. **Maniatis, T., E. F. Frisch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
25. **Marlor, R. L., S. M. Parkhurst, and V. G. Corces.** 1986. The *Drosophila melanogaster* gypsy transposable element encodes putative gene products homologous to retroviral proteins. Mol. Cell. Biol. **6**:1129–1134.
26. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61**:3020–3029.
27. **Moore, R., M. Dixon, R. Smith, G. Peters, and C. Dickson.** 1987. Complete nucleotide sequence of a milk-transmitted mouse mammary tumor virus: two frameshift suppression events are required for translation of *gag* and *pol*. J. Virol. **61**:480–490.
28. **Mount, S. M., and G. M. Rubin.** 1985. Complete nucleotide sequence of the *Drosophila* transposable element copia: homology between copia and retroviral proteins. Mol. Cell. Biol. **5**:1630–1638.
29. **Ono, M., and H. Ohishi.** 1983. Long terminal repeat sequences of intracisternal A-particle genes in the Syrian hamster genome: identification of $tRNA^{phe}$ as the putative primer tRNA. Nucleic Acids Res. **11**:7169–7179.
30. **Ono, M., H. Toh, T. Miyata, and T. Awaya.** 1985. Nucleotide sequence of the Syrian hamster intracisternal A-particle gene: close evolutionary relationship of type A particle gene to types B and D oncovirus genes. J. Virol. **55**:387–394.
31. **Perk, K., and J. E. Dahlberg.** 1974. Murine intracisternal A-type particles fail to separate from the membrane of the endoplasmic reticulum. J. Virol. **14**:1304–1306.

31a. **Reuss, F. U.** Unpublished data.

32. **Saigo, K., W. Kugimiya, Y. Matsuo, S. Inouye, K. Yoshioka, and S. Yuki.** 1984. Identification of the coding sequence for a reverse transcriptase-like enzyme in a transposable genetic element in *Drosophila melanogaster*. Nature (London) **312**:659–661.
33. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
34. **Schaller, H. C., S. Druffel-Augustin, and S. Dübel.** 1989. Head activator acts as an autocrine growth factor for NH15-CA2 cells in the G_2/mitosis transition. EMBO J. **8**:3311–3318.
35. **Shiba, T., and K. Saigo.** 1983. Retrovirus-like particles containing RNA homologous to the transposable element copia in *Drosophila melanogaster*. Nature (London) **302**:119–124.
36. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–546.
37. **Sonigo, P., C. Barker, E. Hunter, and S. Wain-Hobson.** 1986. Nucleotide sequence of Mason-Pfitzer monkey virus: an immunosuppressive D-type retrovirus. Cell **45**:375–385.
38. **Varmus, H.** 1988. Retroviruses. Science **240**:1427–1435.
39. **von Heijne, G.** 1986. A new method for predicting signal sequence cleavage sites. Nucleic Acids Res. **4**:4683–4690.