Nucleotide Sequence Analysis and Enhancer Function of Long Terminal Repeats Associated with an Endogenous African Green Monkey Retroviral DNA

MICHAEL KESSEL^{1†} AND ARIFA S. KHAN^{2*}

Laboratory of Molecular Virology, Division of Cancer Etiology, National Cancer Institute,¹ and the Laboratory of Molecular Microbiology, National Institute of Allergy and Infectious Diseases,² Bethesda, Maryland 20205

Received 17 December 1984/Accepted 22 March 1985

The nucleotide sequence and enhancer activity of the long terminal repeats (LTRs) associated with a cloned endogenous African green monkey (AGM) retroviral DNA designated as λ -AGM-1 was studied. A unique feature of the endogenous AGM proviral LTRs was the presence of multiple copies of two types of directly repeating units in the U₃ region: 16 8-base-pair (bp) repeats were present in the 5' LTR and 12 were present in the 3' LTR which were bound by a 6-bp perfect direct repeat; tandem duplication of a 32-bp sequence resulted in 3.5 copies in the 5' LTR and 2.5 copies in the 3' LTR. Nucleotide sequence homology was seen between the 8-bp direct repeats located in the AGM proviral LTRs and a 10-bp repeat unit of the deca-satellite present in AGM cellular DNA. The 32-bp repeats of the AGM proviral LTRs contained sequences which were related to the SV40 21-bp repeats and to the "core" of the SV40 72-bp enhancer element. Furthermore, the AGM provirus was distinct from known infectious retroviruses due to the presence of a primer-binding sequence complementary to the 3' terminus of mammalian tRNA^{Gly}. Functional analysis of the 3' LTR present in λ -AGM-1 DNA by chloramphenicol acetyltransferase assay demonstrated enhancer activity associated with the 32-bp direct repeats. Sequences outside the 32-bp unit were necessary for full activator function, suggesting the presence of multiple enhancer domains in the AGM provirus.

Retroviral-related sequences have been shown to exist as an integral part of the chromosomal DNA of a large variety of mammalian species (5). The majority of the endogenous proviruses are defective for virus production. In the murine system, although about 30 to 50 copies of murine leukemia virus (MuLV)-related DNAs are present per haploid genome (1, 3, 10, 36), there are only a few virus-inducible loci (18, 30). A large number of defective endogenous MuLV DNAs play a role in formation of recombinant MuLVs (16). The long terminal repeats (LTRs) of both mammalian (19) and avian (21) retroviruses contain enhancer elements (12). In some cases, the LTRs associated with exogenous viruses activate transcription of neighboring cellular genes subsequent to their integration into chromosomal DNA, resulting in malignant disease (14). It thus seems possible that some of the endogenous proviral LTRs could function as cellular enhancers and regulate the expression of proviral or cellular genes located in their vicinity.

Cellular enhancer sequences have previously been identified by using simian virus 40 (SV40) (6) or BK virus (28) enhancers as probes to screen a human genomic DNA library in lambda phage. The SV40- or BK virus-related cellular elements had a low potential for enhancing transcription as compared with their viral counterparts. In this paper, we have approached the study of cellular enhancers by analyzing the structure and function of the LTRs associated with an endogenous MuLV-related retrovirus previously isolated from African green monkey (AGM) DNA and designated as λ -AGM-1 (23). Based upon nucleic acid hybridization and restriction map analysis, λ -AGM-1 DNA was

closely related to the baboon endogenous virus DNA in the *pol* region (24). Here, we report nucleotide sequence analysis of the LTRs and associated regions present in an endogenous AGM provirus and a study of the endogenous LTR enhancer function. The endogenous AGM provirus was distinct from other known infectious retroviruses due to its unique LTR structure and primer-binding site (PBS). The 3' AGM proviral LTR exhibited enhancer activity in the chloramphenicol acetyltransferase (CAT) assay that could be attributed to a 32-bp repeating unit. This activity was 5 to 10-fold less than the enhancement observed in the presence of the SV40 viral enhancer.

MATERIALS AND METHODS

Recombinant DNA clones. λ -AGM-1 proviral DNA was previously cloned from a partial *Eco*RI AGM genome library in λ Charon 4A by using MuLV DNA segments as probes under low-stringency hybridization conditions (23). The cloned DNA segment was 17 kilobases (kb) in size and appeared to contain an intact endogenous provirus with repeated sequences at its 5' and 3' termini located about 8.4 kb apart. DNA segments containing the terminal repeats were subcloned into M13mp8 and M13mp9 replicative-form vector DNAs (25) for nucleotide sequence analysis and for use as probes (see Fig. 1) and into pA10cat2 vector DNA (19) for analysis of CAT gene expression (see Fig. 4).

Nucleotide sequencing. *Escherichia coli* JM103 cells were transformed with recombinant M13 replicative-form DNAs, and phage containing a DNA insert (white plaques) was isolated. Single-stranded phage DNA was prepared as described elsewhere (31), and its nucleotide sequence was determined by the dideoxy chain termination method of Sanger et al. (32).

Biochemicals. Enzymes were purchased from New England Biolabs, Beverly, Mass., or Boehringer Mannheim,

* Corresponding author.

† Present address: Institut für Microbiologie und Zentrum für Molekulare Biologie, 6900 Heidelberg, Im Neuenheimer Feld 364, Federal Republic of Germany.

FIG. 1. Structure of cloned λ -AGM-1 DNA. The retroviral genome organization is shown at the top of the figure. The restriction sites indicated correspond to those which were used in sequencing or for obtaining subgenomic DNA probes. Restriction enzyme abbreviations: A, *Ava*I; E, *Eco*RI; H, *Hind*III; R, *Rsa*I; B, *Bam*HI; S, *Sau*3A. The restriction map for *Ava*I, *Rsa*I, and *Sau*3A is incomplete. Symbols: $\sim$, cellular DNA flanking the provirus; $\blacksquare$, regions from which the probes were derived. Arrows indicate extent and direction of sequencing.

Indianapolis, Ind., and used in buffer conditions specified by the manufacturer; synthetic *Bgl*II linkers were purchased from Collaborative Research, Waltham, Mass.; acetyl coenzyme A was purchased from P-L Biochemicals, Milwaukee, Wis.; [14 C]chloramphenicol was from New England Nuclear Corp., Boston, Mass.; and [α - 32 P]ATP was purchased from Amersham Corp., Arlington Heights, Ill.

Southern blot analysis. High-molecular-weight AGM liver DNA was prepared from fresh liver tissue as previously described (2). Cloned and total-cell DNAs were cleaved with restriction enzymes and electrophoresed in 1.0 and 1.4% agarose gels at 40 V overnight and transferred to nitrocellulose membranes (34). Blot hybridization experiments were carried out under standard conditions as previously described (15).

Transfection and CAT assay. The AGM kidney-derived cell line CV-1 was used in this study. Cells (8×10^5 to 1×10^6) were plated 24 h before the transfection. The transfection procedure for enzymatic activity of CAT has previously been described in detail (11, 19). In a standard reaction, two-thirds (i.e., 100 μ l) of the cell extract prepared from cells on a 10-cm tissue culture dish was incubated for 60 min at 0.5 mM acetyl coenzyme A and 1 to 10 μ l of [14 C]chloramphenicol (50 mCi/mmol). The radioactive reaction products were separated on silica gel thin-layer plates and autoradiographed.

Construction of CAT plasmids containing AGM LTR sequences. The following recombinant CAT plasmids were used in this study (see Fig. 4).

(i) **pAGM-1.** A 1.85-kb *Bam*HI fragment from λ -AGM-1 was inserted in the sense orientation into a *Bgl*II site located at the 5' end of the recipient vector pA10cat2 (19). This segment extended from a *Bam*HI site located in the *env* region at a position 6.65 kb from the 5' end of the provirus in λ -AGM-1 DNA to a *Bam*HI site located 96 bp into the 3' cellular flanking region and contained the intact 3' AGM proviral LTR.

(ii) **pAGM-2.** A 5' deletion encompassing 300 bp of *env* sequences was introduced into pAGM-1 by cutting with *Ava*I and religating the vector without the small *Ava*I *env* DNA fragment.

(iii) **pAGM-3.** pA10cat2 was cleaved with *Sal*I and *Hind*III, the 5' overhangs were filled in with the large fragment of DNA polymerase I (Klenow fragment), and the isolated vector was blunt-end ligated, generating a new *Hind*III site. A *Hind*III fragment from pAGM-1 beginning with the *Hind*III

site at position 138 in the 3' LTR and extending into 3' flanking cell DNA was inserted in the sense orientation into the new *Hind*III site in pA10cat2. This fragment contained the SV40 early transcriptional unit and ended with the *Hind*III site located at the 5' end of the CAT gene in pAGM-1.

(iv) **pAGM-4.** An *Rsa*I fragment containing 438 bp of the 3' LTR including the U₃ region and 91 bp of adjacent *env* sequences was isolated from pAGM-1. This segment was inserted into the *Bgl*II site of pA10cat2 in the sense orientation by means of *Bgl*II linkers.

(v) **pAGM-5 and pAGM-6.** A 270-bp *Hind*III-*Rsa*I fragment extending from LTR position 140 to 438 was isolated, and its ends were converted to *Bgl*II sites by the addition of synthetic linkers. Insertion into the *Bgl*II site of pA10cat2 resulted in pAGM-5 (sense orientation) and pAGM-6 (antisense orientation).

(vi) **pAGM-7 and pAGM-8.** The same 270-bp *Hind*III-*Rsa*I fragment as described above was inserted 3' of the transcriptional unit into the *Bam*HI site of pA10cat2 to generate pAGM-7 (sense) and pAGM-8 (antisense).

RESULTS

Nucleotide sequence analysis of the LTRs associated with cloned AGM proviral DNA. The nucleotide sequences of the 5' and 3' LTRs present in cloned λ -AGM-1 DNA (23) were determined by using LTR-containing recombinant M13 DNAs and following the sequencing strategy shown in Fig. 1. The results of this analysis are presented in Fig. 2. The extent of the LTR structure was determined by comparing nucleotide sequences from the 5' and 3' regions of the provirus. As with all known integrated retroviral DNAs, the AGM proviral LTR begins with TG and ends with CA (38, 40). In fact, these dinucleotides compose the inverted repeats in the AGM LTR which characterize the cell-virus junction. The AGM proviral LTR contains the transcriptionally important regulatory signals such as the CCAAT box (positions 376 through 380), a Goldberg-Hogness sequence (positions 425 through 430), a polyadenylic acid signal (positions 495 through 500) and a polyadenylic acid addition site (positions 525 through 526) (Fig. 2). A unique feature of the LTRs associated with the AGM provirus is the presence of two types of repeating structures. Tandem duplication of a 32-bp sequence (positions 162 through 193) located at approximately the same position relative to the CCAAT box as the 72-bp direct repeat in the MSV LTR (7) resulted in 3.5

FIG. 2. Comparison of nucleotide sequences in the 5' and 3' LTRs associated with λ -AGM-1 proviral DNA. Sequences which compose the LTR structure are located within the brackets and numbered from the 5' end of the 5' LTR. Nucleotides which are identical in the 5' and 3' LTRs are represented by dots, gaps indicate nucleotides which are absent, nucleotide differences are shown, and *Hind*III and *Rsa*I restriction enzyme sites are indicated. The transcriptionally important regulatory signals are underlined, 32-bp direct repeats are indicated by heavy overhead arrows and lettered A through D, 8-bp direct repeats are indicated by thin overhead arrows and numbered 1 through 16, and the direct repeat sequence flanking the 8-bp repeating units is boxed. Cellular sequences which are duplicated at the host-provirus junction are located outside the brackets. Sequences which compose the polypurine tract and the PBS are also shown.

copies in the 5' LTR and 2.5 copies in the 3' LTR. Starting at position 17 from the 5' end of the LTR, 16 copies of an 8-bp sequence were identified in the 5' LTR, and 12 copies were identified in the 3' LTR. The short 8-bp tandem repeating units were flanked by a 6-bp perfect direct repeat (positions 10 through 15 and 150 through 155). The consensus sequence of the 8-bp repeat units is 5' A $\overline{A}$ ACTTAG 3'.

Nucleotide sequence comparison of the 5' and 3' LTRs indicated that the major difference between the two LTRs was due to the absence of four 8-bp repeats and one 32-bp repeat in the 3' LTR. Other changes in the 3' LTR included a single-base insertion at position 578 in the 3' LTR, a cluster of five base changes where the sequence 5'-GTGCG-3' in the 5' LTR was replaced by the complementary sequences 5'-CACGC-3' in the 3' LTR (positions 319 through 323), and five random single-base substitutions.

The nucleotide sequence of the AGM proviral LTR showed little homology with any known retroviral LTR. A region of 21 bp (starting at position 296 in the AGM LTR) bears sequence homology (17 of 19 nucleotides) to sequences located from positions 54 through 72 in the U₃ region of the human adult T-cell leukemia virus (33). A 13-bp nucleotide stretch also started at position 296 in the AGM LTR was

closely related (12 of 13 nucleotides) to sequences extending from positions 156 through 165 in the baboon endogenous virus LTR (37). Sequences which loosely resemble motifs of the SV40 transcriptional control region were present upstream of the CCAAT box. The sequence CCGCCC, which is repeated twice in each of the three SV40 21-bp repeats (8), occurred three times (starting at positions 172, 204, and 236) in the 5' AGM proviral LTR and once (starting at position 172) in the 3' LTR. A copy of the 32-bp repeat unit present in AGM proviral LTR contained a sequence TGTGGCTA, located at positions 251 through 258, which may be related to the core sequence of the SV40 72-bp enhancer element, TGTGG(AAA)G (12). Nucleotide sequence homology was seen between the 8-bp direct repeats (AAACTTAG) located in the U₃ region of the AGM LTR and a 10-bp repeat unit (AAACTAGGGT) present in deca-satellite sequences of the AGM genome (22).

Sequences juxtaposing the LTRs contained structural features characteristic of retroviruses. These included a duplication of cellular sequences located at the right and left cell-virus junctions (38, 40). However, the 7-bp duplication was not a perfect direct repeat: the 7 bases on the 5' side (TTAGGAG) differed in two nucleotides from those on the

FIG. 3. Southern blot analysis of AGM genomic DNA. Cloned λ -AGM-1 DNA (lanes 1, 3, 5, and 7) and cellular AGM liver DNA (lanes 2, 4, 6, and 8) were cleaved with *Bam*HI (A) or *Hind*III (B) restriction enzymes and electrophoresed at 40 V overnight on 1.0 or 1.4% agarose gels, respectively. The DNA fragments were subjected to Southern blot analysis and probed with the *Hind*III LTR-specific (lanes 1, 2, 5, and 6) or *Rsa*I *env*-specific (lanes 3, 4, 7, and 8) DNA probes as described in the text. The sizes (in kb) of the fragments present in the cloned AGM proviral DNA which annealed to the probes are shown. (C) Diagram indicating the *Hind*III fragments in λ -AGM-1 DNA which hybridize to the LTR and *env*-specific DNA probes. Designations are the same as in Fig. 1. Symbols: $\rightarrow$, 8-bp repeats; $\blacktriangleright$, 32-bp repeats.

3' side (TTAACAG). The polypurine tract preceding the 3' LTR contained sequences which are unrelated to the consensus sequence of the polypurine regions of known mammalian retroviruses (4). Viral sequences extending from the 5' LTR into the *gag* region contained a PBS adjacent to the 5' LTR. Unlike known infectious type C mammalian or avian retroviruses which contain PBS sequences complementary to tRNA^{Pro} or tRNA^{Trp}, respectively (4), sequences in the PBS region of the cloned AGM proviral DNA were complementary in 20 of 21 bases to the 3'-terminal sequence of human placenta tRNA^{Gly} (13).

Southern blot analysis of AGM genomic DNA. To analyze the distribution of λ -AGM-1 DNA-related sequences in the AGM genome, total liver DNA was cleaved with *Bam*HI and subjected to Southern blot analysis (Fig. 3A) with DNA probes derived from either the LTR (lanes 1 and 2) or *env* (lanes 3 and 4) regions. The *Hind*III LTR probe consisted of a 0.25-kb DNA segment encompassing nucleotides 139 through 394 in the 5' LTR including the 32-bp direct repeats. After hybridization to *Bam*HI-cleaved λ -AGM-1 DNA (Fig. 3A, lane 1), this LTR probe annealed with the 8.8-kb DNA segment containing the 5' LTR and the 1.85-kb fragment

encompassing the 3' LTR. The map positions of the *Bam*HI sites in the provirus are indicated in Fig. 1. The *Rsa*I *env*-specific DNA probe, which extended from 8.0 to 8.15 kb from the 5' end of the provirus, hybridized to only the 1.85-kb *Bam*HI DNA segment (Fig. 3A, lane 3) located at the 3' end of the λ -AGM-1 DNA. The same labeled probe hybridized to at least 20 different fragments in *Bam*HI-restricted AGM cellular DNA (Fig. 3A, lane 4). These reactive cleavage products represented 3' virus-cell junction fragments. These segments also hybridized to the LTR-specific DNA probe (Fig. 3A, lane 2). These results suggest that the majority of LTRs present in the AGM DNA are associated with proviruses.

A unique feature of the LTRs associated with λ -AGM-1 DNA was extensive duplication of an 8- and a 32-bp sequence in the U₃ region (Fig. 2). If the presence of these multiple copies of direct repeats in λ -AGM-1 proviral LTRs was an artifact of the cloning technique, then such duplications should not exist in the LTRs associated with proviral DNAs present in the AGM genome. To analyze the nature of the proviral LTRs in AGM genomic DNA, Southern blot analysis of *Hind*III-digested DNAs was performed with

FIG. 4. Schematic representation of the AGM endogenous proviral LTR and indication of subcloned fragments in pA10cat2 DNA. Some of the structural features of the 3' LTR associated with λ -AGM-1 DNA are indicated diagrammatically in panel I. The restriction fragments containing LTR sequences which were subcloned into the vector DNA pA10cat2 to generate recombinant plasmids pAGM-1 through pAGM-8 are shown. The insertion site with respect to the CAT gene in pA10cat2 is indicated in panel II. Symbols: $\rightarrow$, inverted repeats; $\blacktriangleright$, 8-bp direct repeats; $\blacktriangleleft$, 32-bp direct repeats; $\blacktriangleright$, direction of SV40 early transcription; $\blacksquare$, SV40 early promoter region; ∇ , the insertion site of LTR-containing segments with respect to the CAT gene in recombinant plasmids pAGM-1 through pAGM-8. Restriction site abbreviations: H, *Hind*III; R, *Rsa*I; B, *Bam*HI; A, *Ava*I; Bg, *Bgl*II; Sph, *Sph*I; S, *Sal*I.

32 P-labeled LTR- or *env*-specific DNA probes. There is at least one *Hind*III site in both the 5' and 3' LTRs associated with cloned λ -AGM-1 proviral DNA (Fig. 3C). The location of this *Hind*III site (nucleotide position 138) between the 8- and 32-bp direct repeat sets enabled us to study duplications of the 32-bp sequence by using the *Hind*III LTR DNA probe which hybridized to the 0.25- and 6.3-kb *Hind*III fragments containing the 32-bp repeats in λ -AGM-1 (Fig. 3B, lane 5). The 8-bp sequence was contained within the 0.8-kb *Hind*III fragment which hybridized to the *Rsa*I *env*-specific DNA probe (Fig. 3B, lane 7). The location of the reactive segments is indicated in Fig. 3C. Fragments were generated from endogenous proviruses present in the AGM genome which corresponded to the LTR- or *env*-reactive DNA fragments present in the cloned provirus. Heterogeneity was observed in the size of the 0.25-kb LTR-reactive *Hind*III fragment containing the 32-bp direct repeats and in the size of the 0.8-kb *env*-reactive segment containing the 8-bp repeats (Fig. 3B, lanes 6 and 8, respectively). This variation in fragment size may reflect differences in the number of repeating units in the proviral LTR. The presence of *Hind*III fragments in the AGM DNA which were greater in size than the 0.25-kb LTR-reactive (Fig. 3B, lane 6) and 0.8-kb *env*-reactive (lane 8) segments present in λ -AGM-1 DNA suggests restriction site polymorphism in the other AGM endogenous proviruses. Since sequences in the 8-bp repeat units of the LTR are highly related to the *Hind*III cleavage site, nucleotide changes within this short tandem repeat structure may generate 5' and 3' junction fragments.

Study of the enhancer activity of endogenous AGM proviral LTRs. To determine whether the LTRs associated with λ -AGM-1 DNA contained enhancer function, fragments

encompassing segments of the 3' LTR were subcloned into the vector pA10cat2 which lacks SV40 enhancer sequences but contains SV40 promoter elements (19) and assayed for CAT gene expression (11). The CAT assay has been used as a convenient and sensitive method for detecting enhancer function. The recombinant plasmid constructs used in the analysis are shown in Fig. 4 and their derivation is described above. pAGM-1 consists of a 1.85-kb DNA segment from λ -AGM-1 which contains about 1.5 kb of 3' *env* sequences, the intact 3' AGM proviral LTR, and 96 bp of the 3' flanking cellular DNA. In pAGM-2, the *env* sequences were deleted up to the *Ava*I site, leaving 1,500 bp of AGM DNA. pAGM-3 represents a further 5' deletion which goes up to the *Hind*III site located at position 138 in the LTR and encompasses sequences in the 5' portion of the LTR including the 8-bp repeats. The *Rsa*I segment in pAGM-4 contains 100 nucleotides from the 3' terminus of the *env* region of U₃ sequences of the LTR, including both the 8- and 32-bp repeating units: the entire R and U₅ portions of the 3' LTR have been deleted. The 270-bp *Hind*III-*Rsa*I fragment contains the 32-bp repeats but not the 8-bp repeats and is present in the sense and antisense orientations in plasmids pAGM-5 and pAGM-6, respectively. The constructs pAGM-1 through pAGM-6 contain 3' LTR sequences with intact CCAAT and TATA boxes. In recombinant plasmid DNAs pAGM-1 through pAGM-6, these sequences were cloned upstream of the SV40 promoter and CAT gene (Fig. 4, panel II).

Upon transfection of CV-1 cells with pAGM-1 through pAGM-6 recombinant plasmid DNAs, CAT gene expression was enhanced by a factor of 20 to 40 as compared with the enhancer-free vector pA10cat2 (Fig. 5). This activity was 5 to 10-fold less than the enhancement observed in the pres-

FIG. 5. CAT assay of pAGM plasmid DNAs. Thin-layer separation of the reaction products on silica gel plates is shown. The indicated plasmids were transfected onto CV-1 cells, cell extract was prepared after 48 h, the extract equivalent of two-thirds of a 10-cm tissue culture dish (100 μ l) was incubated for 60 min as described in the text, and radioactive material was extracted with ethyl acetate and separated. The upper two spots represent acetylated chloramphenicol, and the lower spot represents unconverted chloramphenicol. The percent conversion represents the percentage of [14 C]chloramphenicol converted to an acetylated form. This was determined by measuring the radioactivity of spots scraped from the plates in a liquid scintillation counter.

ence of the SV40 enhancer (data not shown). Analysis of the various constructs suggested that the enhancer activity could be attributed to the 32-bp repeating units in the LTR. The enhancer activity was maintained in pAGM-6, in which the 270-bp *HindIII-RsaI* segment containing the 32-bp repeats was cloned in the antisense orientation at the 5' end of the CAT gene. Insertion of this *HindIII-RsaI* segment in the sense orientation at the 3' terminus of the transcriptional unit, however, resulted in little or no enhancement of expression (pAGM-7), whereas a reduced level of enhancement was observed with this fragment in the antisense orientation at the 3' location (Fig. 5, pAGM-8). It should be noted that viral sequences outside the 270-bp segment which contains the 32-bp repeats were necessary for full enhancing activity; a higher percent conversion was seen with pAGM-2 DNA than with pAGM-5 DNA.

DISCUSSION

We have analyzed the structure and function of the LTRs associated with an endogenous MuLV-related provirus cloned previously under nonstringent hybridization conditions from an AGM genomic DNA library (23). The endogenous AGM proviral DNA, designated as λ -AGM-1, was most closely related in the *pol* region to the baboon endogenous virus, based upon nucleic acid hybridization experiments and the conservation of 8 of 10 restriction enzyme sites between the two retroviral DNAs (24). Southern blot analysis demonstrated the presence of at least 20 copies of λ -AGM-1-related sequences present in AGM DNA. In each case, homologous, LTR-like sequences were associated with *env*-specific regions of DNA. However, no infectious virus has been induced from AGM cells (39).

Although the cloned DNA in λ -AGM-1 was structurally related to a retrovirus, sequences in its LTRs and associated regions could be distinguished from those present in known infectious retroviruses. Reverse transcription of an infectious retroviral genome leads to the formation of two virtually identical LTR units at the 5' and 3' termini of the provirus (40). Integration of proviral DNA into the host genome results in perfect direct duplication of cellular sequences at the site of insertion. Both of these features, which are characteristic of recently integrated proviruses, were absent in the endogenous AGM proviral DNA: nucleotide differences were seen between the endogenous 5' and 3' LTRs; two mismatched bases were present in the AGM host direct repeat. These results suggest that the endogenous AGM retroviral DNA may represent an ancient provirus. The inability to form infectious virus might be due to nucleotide changes resulting in deletions, insertions, and point mutations in the LTR as well as other regions of the viral genome. To date, all infectious retroviruses require tRNA^{Pro}, tRNA^{Trp}, or tRNA^{Lys} for initiation of minus strand synthesis (4). In contrast, the PBS in the AGM provirus was shown to be complementary to the 3' terminus of mammalian tRNA^{Gly} (13). It has recently been shown that the retrovirus-like 30S (VL30) mouse genetic element, which is incapable of forming virus particles, uses tRNA^{Gly} as the minus strand primer (26, 29). Furthermore, an endogenous murine (27) and an endogenous human (35) retrovirus have been shown to contain sequences complementary to the 3' ends of tRNA^{Gln} or tRNA^{Glu}, respectively (or a related species of tRNA). A novel PBS different from that present in infectious retroviruses seems to be a characteristic of endogenous, non-virus-forming, retroviral DNAs.

A unique feature of the endogenous AGM proviral LTRs was the presence of multiple tandem direct repeats in the U₃ region. An 8-bp sequence located 17 nucleotides from the 5' end of the LTR unit was present in 16 copies in the 5' LTR and 12 copies in the 3' LTR. Nucleotide sequence analysis showed similarity between the 8-bp repeats (AAACTTAG) and a repeat unit contained in the deca-satellite (AAACTAGGGT) present in AGM cellular DNA as described by Maresca and Singer (22). They have reported that extensive variation of the consensus sequence (AAACCGGNTC) is seen in the 10-bp repeats of the deca-satellite. The relatedness of the 8-bp repeating unit present in AGM proviral LTRs to cellular repeats in the AGM genome, combined with the presence of a perfect direct duplication abutting the 8-bp repeats, suggests insertion of a deca-satellite-related cellular sequence into the AGM proviral LTR. Insertion of a cellular repetitive DNA segment into the U₃ region of an endogenous murine retroviral LTR has previously been reported (17). The role of such cellular insertions into retroviral LTRs is unclear.

The 5' AGM proviral LTR contained 3.5 copies of a 32-bp sequence located in a position analogous to that of the enhancer element in the MSV LTR (7); 2.5 copies were present in the 3' LTR. The sequence CCGCCC, which is characteristic of the SV40 21-bp repeats (8), was identified in the 32-bp repeating unit in the AGM LTR. Previous studies with SV40 deletion mutants have shown that the copy number of the short GC-rich repeats determines the efficiency of the early promoter region (8, 9). Reiteration of the 32-bp unit in the AGM proviral LTRs resulted in three copies of the CCGCCC sequence in the 5' LTR and one copy in the 3' LTR. Duplication of sequences has been reported to generate enhancers more active than the original. Strong activators have been created in the cases of SV40 (41) and Rous sarcoma virus (20) as a result of tandem duplication of nonenhancing sequences. The presence of multiple tandem repeats of the 8- and 32-bp sequences may be significant in enhancer or promoter function or both of the endogenous AGM proviral LTRs.

A functional analysis by CAT assay demonstrated enhancer activity associated with the 32-bp tandem direct repeats. Criteria for an enhancer function were fulfilled in terms of orientation independence. Activation of the CAT gene by the 32-bp repeats located at its 3' end, which has been postulated for classical enhancers, was only observed at a low level in the antisense orientation. A similar decrease in activity has previously been reported for the murine sarcoma virus enhancer in experiments with the same parental vector (19). The decrease in enhancer activity may be due to availability of alternate promoters in the plasmid constructs. The presence of the 32-bp repeats enhanced the expression of the CAT gene 20- to 40-fold as compared with the enhancer-free vector pA10cat2. This, however, was 5- to 10-fold less than the activity achieved with the 72-bp repeats from SV40. Retroviral sequences outside the LTR have previously been shown to contribute to expression of full enhancing activity (20, 21). The enhancer activity of the 32-bp repeats in the AGM LTR was augmented by the presence of other neighboring viral sequences, including the 8-bp repeats, thus suggesting the presence of multiple enhancer domains in the AGM proviral DNA.

The role of endogenous retrovirus-like sequences resident in higher eucaryotic cells remains obscure. Nevertheless, the presence of transcriptional control elements such as enhancers in proviral genomes suggests that they may have the ability to activate otherwise silent cellular genes.

ACKNOWLEDGMENTS

We thank Peter Gruss, George Khoury, and Malcolm A. Martin for initiating the study, discussion, and review of the manuscript, Willard W. Davis for excellent technical assistance, and Laura A. Wilders and Susan K. Rosenfeld for typing the manuscript.

LITERATURE CITED

- Callahan, R., R. E. Benveniste, M. M. Lieber, and G. J. Todaro. 1974. Nucleic acid homology of murine type-C viral genes. *J. Virol.* 14:1394-1403.
- Chan, H. W., T. Bryan, J. L. Moore, S. P. Staal, W. P. Rowe, and M. A. Martin. 1980. Identification of ecotropic proviral sequences in inbred mouse strains with a cloned subgenomic DNA fragment. *Proc. Natl. Acad. Sci. U.S.A.* 77:5779-5783.
- Chattopadhyay, S. K., D. R. Lowy, N. M. Teich, A. S. Levine, and W. P. Rowe. 1974. Qualitative and quantitative studies of AKR-type murine leukemia virus sequences in mouse DNA. *Cold Spring Harbor Symp. Quant. Biol.* 39:1085-1101.
- Chen, H. R., and W. C. Barker. 1984. Nucleotide sequences of the retroviral long terminal repeats and their adjacent regions. *Nucleic Acids Res.* 12:1767-1778.
- Coffin, J. 1982. Endogenous retroviruses, p. 1109-1203. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), *Molecular biology of tumor viruses: RNA tumor viruses*, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- Conrad, S. E., and M. R. Botchan. 1982. Isolation and characterization of human DNA fragments with nucleotide sequence homologies with the simian virus 40 regulatory region. *Mol. Cell. Biol.* 2:949-965.
- Dhar, R., W. L. Clements, W. L. Enquist, and G. F. Vande Woude. 1980. Nucleotide sequence of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. *Proc. Natl. Acad. Sci. U.S.A.* 77:3937-3941.
- Everett, R. D., D. Boty, and P. Chambon. 1983. The repeated GC-rich motifs upstream from the TATA box are important elements of the SV40 early promoter. *Nucleic Acids Res.* 11:2447-2464.
- Fromm, M., and P. Berg. 1983. Transcription in vivo from SV40 early promoter deletions mutants without repression by large T antigen. *J. Mol. Appl. Genet.* 2:127-135.
- Gelb, L. D., J. B. Milstien, M. A. Martin, and S. A. Aaronson. 1973. Characterization of murine leukemia virus-specific DNA present in normal mouse cells. *Nature (London)* 244:76-79.
- Gorman, G. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044-1051.
- Gruss, P. 1984. Magic enhancers? *DNA* 3:1-5.
- Gupta, R. C., B. A. Roe, and K. Randerath. 1979. The nucleotide sequence of human tRNA^{Gly} (anticodon GCC). *Nucleic Acids Res.* 7:959-970.
- Hayward, W. S., B. G. Neel, and S. M. Astrin. 1981. Activation of a cellular *onc* gene by promoter insertion in ALV-induced lymphoid leukemia. *Nature (London)* 290:475-480.
- Israel, M. A., D. F. Vanderryn, M. L. Meltzer, and M. A. Martin. 1980. Characterization of polyoma viral DNA sequences in polyoma-induced hamster tumor cell lines. *J. Biol. Chem.* 255:3798-3805.
- Khan, A. S. 1984. Nucleotide sequence analysis establishes the role of endogenous murine leukemia viral DNA segments in formation of recombinant mink cell focus-forming murine leukemia viruses. *J. Virol.* 50:864-871.
- Khan, A. S., and M. A. Martin. 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. *Proc. Natl. Acad. Sci. U.S.A.* 80:2699-2703.
- Kozak, C. A., and W. P. Rowe. 1980. Genetic mapping of xenotropic murine leukemia virus-inducing loci in five mouse strains. *J. Exp. Med.* 152:219-228.
- Laimins, L. A., P. Gruss, R. Pozzati, and G. Khoury. 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. *J. Virol.* 49:183-189.
- Laimins, L. A., P. Tschlis, and G. Khoury. 1984. Multiple

- enhancer domains in the 3' terminus of the Prague strain of Rous sarcoma virus. *Nucleic Acids Res.* **12**:6427-6442.
21. Leciw, P. A., J. M. Bishop, H. E. Varmus, and M. R. Capecchi. 1983. Location and function of retroviral and SV40 sequences that enhance biochemical transformation after microinjection of DNA. *Cell* **33**:705-716.
 22. Maresca, A., and M. F. Singer. 1983. Deca-satellite: a highly polymorphic satellite that joins α -satellite in the African green monkey genome. *J. Mol. Biol.* **164**:493-511.
 23. Martin, M. A., T. Bryan, T. F. McCutchan, and H. W. Chan. 1981. Detection and cloning of murine leukemia virus-related sequences from African green monkey liver DNA. *J. Virol.* **39**:835-844.
 24. 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. U.S.A.* **78**:4892-4896.
 25. Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**:269-276.
 26. Norton, J. D., J. Conner, and R. J. Avery. 1984. Unusual long terminal repeat sequence of a retrovirus transmissible mouse (VL30) genetic elements: identification of function domains. *Nucleic Acids Res.* **12**:3445-3460.
 27. Ou, C. Y., L. R. Boone, and W. K. Yang. 1983. A novel sequence segment and other nucleotide structural features in the long terminal repeat of a BALB/c mouse genomic leukemia virus-related DNA clone. *Nucleic Acids Res.* **11**:5603-5620.
 28. Rosenthal, N., M. Kress, P. Gruss, and G. Khoury. 1983. The BK viral enhancer element and a human cellular homology. *Science* **222**:749-755.
 29. Rotinan, G., A. Itin, and E. Keshet. 1984. "Solo" large terminal repeats (LTR) of an endogenous retrovirus like gene family (VL30) in the mouse genome. *Nucleic Acids Res.* **12**:2273-2282.
 30. Rowe, W. P. 1978. Leukemia viral genomes in the chromosomal DNA of the mouse. *Harvey Lect.* **71**:173-192.
 31. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161-178.
 32. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5463-5467.
 33. Seiki, M., S. Hatter, and M. Yoshida. 1982. Human adult T-cell leukemia virus: molecular cloning of the provirus DNA and the unique terminal structure. *Proc. Natl. Acad. Sci. U.S.A.* **79**:6899-6902.
 34. Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **38**:503-517.
 35. Steele, P. E., A. B. Rabson, T. Bryan, and M. A. Martin. 1984. Distinctive termini characterize two families of human endogenous retroviral sequences. *Science* **225**:943-947.
 36. Steffen, D., and R. A. Weinberg. 1978. The integrated genome of murine leukemia virus. *Cell* **15**:1003-1010.
 37. Tamura, T., M. Noda, and T. Takono. 1981. Structure of the baboon endogenous virus genome nucleotide sequences of the long terminal repeat. *Nucleic Acids Res.* **9**:6615-6625.
 38. Temin, H. M. 1981. Structure, variation, and synthesis of retrovirus long terminal repeat. *Cell* **27**:1-3.
 39. Todaro, G. J., R. E. Benveniste, S. A. Sherwin, and C. J. Sherr. 1978. Mac-1, a new genetically transmitted type C virus of primates: "low frequency" activation from stump-tail monkey cell cultures. *Cell* **13**:775-782.
 40. Varmus, H. E. 1982. Form and function of retroviral proviruses. *Science* **216**:812-820.
 41. Weber, F., J. de Villiers, and W. Schaffner. 1984. An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell* **36**:983-992.