JOURNAL OF VIROLOGY, Jan. 1981, p. 226-238
0022-538X/81/010226-13$02.00/0

Vol. 37, No. 1

Regulatory and Coding Potential of the Mouse Mammary Tumor Virus Long Terminal Redundancy

LAWRENCE A. DONEHOWER, ANN L. HUANG, AND GORDON L. HAGER*

Laboratory of Tumor Virus Genetics, National Cancer Institute, Bethesda, Maryland 20205

Molecular clones containing the 3′ half of newly integrated mouse mammary tumor virus (MMTV) DNA with adjacent mouse cellular sequences were characterized. In addition, we cloned the long terminal redundancy joint from the unintegrated circular form of MMTV DNA. The entire nucleotide sequence of the integrated and part of the unintegrated terminal redundancy was determined; this allowed us to delineate the boundaries of the MMTV long terminal redundancy, which comprises 1,327 base pairs. The position of possible RNA polymerase II initiation and termination signals corresponded closely to the expected regions of viral RNA initiation and termination specified by current models. The MMTV long terminal redundancy also contained a large open reading frame with sufficient information for a protein of 198 amino acids. Initial comparison of flanking 3′ cellular sequences from three independent integrated clones suggested there was no host sequence specificity in the MMTV integration event. However, specificity of integration with respect to viral sequences was precise.

It is now well established that RNA tumor viruses contain terminal direct repeat sequences at the ends of both the linear viral DNA intermediate and the integrated provirus (18, 19, 39, 42). These sequences arise as a result of copying regions at the 5′ and 3′ ends of the viral RNA twice during reverse transcription. The long terminal redundancy (LTR) of avian tumor viral DNAs encompasses approximately 300 nucleotides, the LTR of mammalian type C viruses encompasses approximately 600 nucleotides, and the LTR of mouse mammary tumor virus (MMTV), a type B virus, encompasses over 1,200 nucleotides. Topological comparison of the integrated viral DNA with the viral RNA genome suggests that the LTR regions play important roles in both the viral integration process and the subsequent transcription of viral RNA. The presence of information from the 3′ end of viral RNA proximal to the 5′ end of viral RNA predicts (Fig. 1, a) that the promoter for transcription initiation, as well as possible regulatory sites, would be encoded in the 3′ portion of the LTR. Initiation of transcription would begin at the 3′-5′ junction to generate a viral RNA with a mature 5′ end. Alternatively, viral transcription could be promoted from adjacent cell sequences, followed by processing to the mature viral RNAs. In a more complex model, one could envision promotion occurring at a virus-encoded site, with flanking cell DNA contributing regulatory regions. Models invoking a role for adjacent cell sequences imply that cellular genome position may be of importance in regulating the expression of viral sequences. Whether retroviral integration involves the recognition of specific sites in the cellular DNA is thus of considerable relevance to the question of regulation of viral expression.

A second prediction (Fig. 1, b) implicit in the topology of the integrated provirus is that regulatory signals controlling the termination of viral RNA and the addition of the polyadenylic acid sequences should also be specified within virus-encoded DNA. The likely location for these signals would be in the short terminal repeat sequences present at both ends of viral RNA.

The postulated presence of initiation and termination signals within the redundant DNA sequences poses two interesting questions. First (Fig. 1, c), how do RNAs initiated in the leftward LTR escape termination at a site which presumably serves such a function in the rightward LTR? Second (Fig. 1, d), does an identical copy of the hypothetical promoter structure in the rightward LTR serve to promote transcription of downstream cellular information?

In an initial effort to address the issues discussed above, we have determined the complete DNA sequence of the MMTV LTR derived from the newly integrated provirus. To precisely define the boundaries of the MMTV terminal redundancy, we have sequenced (i) the terminal redundancy joint of the unintegrated circular form of MMTV and (ii) the virus-host 3′ joints of three independent integration sites of newly integrated MMTV. The sequence data obtained indicate that the MMTV proviral terminal re-

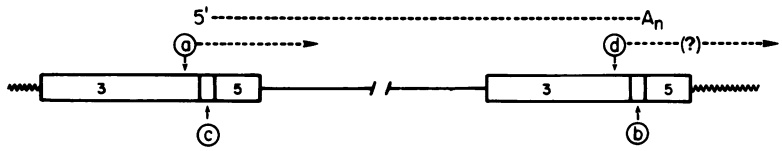


FIG. 1. *Predictions for transcription of integrated retroviral genomes. Viral RNA structure is depicted by the broken line. Below the viral RNA, the integrated form of retroviral DNA is schematically presented, with the terminally repeated DNA sequences indicated by the closed boxes. The double vertical lines within the boxes represent the sequences repeated at each end of genomic RNA; saw-toothed lines indicate host cell DNA. Initiation of transcription is mediated by a hypothetical RNA polymerase II regulatory structure (a) located in the 3' region of the leftward LTR immediately proximal to the 5' segment. Termination of transcription is expected to occur beyond the junction of the 3' and 5' regions of the rightward LTR (b) so as to include the small terminally repeated sequence in the viral RNA. The signal for polyadenylation should occur near the site of termination (b). The model assumes that the newly initiated transcript is able to proceed through the termination signal (c) in the leftward LTR by some undetermined mechanism. Also possible is the downstream transcription of 3' adjacent cellular sequences from the rightward promoter (d).*

dundancy is 1,327 base pairs in length and contains potential RNA promoter and terminator sequences located in positions predicted by the replication model. At the right and left termini of the redundancies are 6-base inverted repeat sequences which form a complimentary palindrome separated by 4 base pairs at the redundancy joint of the unintegrated circle. The sequences of the 3' virus-host DNA joints reveal no common sequence among the cellular flanking sequences, suggesting that the viral integration event is not sequence specific with respect to the cellular DNA. However, the integration event with respect to the viral DNA appears to be precise at the nucleotide level. Unexpectedly, a large open reading frame with a coding potential of 198 amino acids (assuming no introns) is present in the center of the 3'-derived part of the redundancy.

(This work was in partial fulfillment of the requirements of L.D. and A.H. for the Ph.D. degree in genetics at the George Washington University, Washington, D.C.)

MATERIALS AND METHODS

Mammalian cells and virus. Two murine mammary tumor cell lines producing high levels of MMTV (C3H) were utilized. Both were derived from the original C3H MT cell line established by Owens and Hackett (33) from a mammary adenocarcinoma of a C3H/Crgl mouse. The 34I (cl 101) subclone was described by Parks et al. (34); the Mm5mt/c_1 line was developed by Fine et al. (8). MMTV (C3H) virus, harvested from the Mm5mt/c_1 cell supernatant fluid, was obtained from the Intramural Viral Resources Program, National Cancer Institute. Variants of the rat hepatoma (HTC) cell line developed by Thompson et al. (48) were infected with MMTV (C3H) virus; clonal lines containing newly integrated MMTV provirus have been described elsewhere (49).

Mammalian cell culture media. Murine and HTC cell lines were maintained in Dulbecco minimal essential medium containing 10% fetal calf serum, 2 mM L-glutamine, 10 IU of penicillin per ml, and 10 μg of streptomycin per ml.

Bacterial, bacteriophage, and plasmid strains. The EK2-certified lambda phage vector, Charon 4A (3), was supplied by F. Blattner. The EK2 *Escherichia coli* host strain DP50 (*supE supF hsdM hsdR*) was obtained from D. Tiemeier. *E. coli* strain LE392 (*thyA supE supF hsdM*$^+$ *hsdR*) and the EK2 phage vector, λgtWES·λB (26), were supplied by L. Enquist. *E. coli* strain NS428 [N100 (λ*Aam11b2red3cI857Sam7*)] was kindly provided by Nat Sternberg, and *E. coli* strain λdg805 [W3350 (λ*dgal805cI857Sam7*)] was received from F. Blattner. Plasmid pBR322, an EK2-certified vector, was obtained from Bethesda Research Laboratories. The EK2 *E. coli* host χ1776 was generously supplied by Roy Curtiss. EK1 host strain N99 was obtained from A. Das.

Bacterial media and growth. Bacteriophage vectors and recombinants were propagated as described previously (13). *E. coli* χ1776 harboring recombinant plasmids was grown on X broth supplemented with 100 μg of diaminopimelic acid, 50 μg of thymidine, and 20 μg of ampicillin per ml. X broth contains (per liter): 25 g of tryptone, 7.5 g of yeast extract, 20 mM $MgSO_4$, and 50 mM Tris-hydrochloride (pH 7.5). EK1 host *E. coli* N99 harboring Amp^r plasmids was grown on L-broth containing 25 μg of ampicillin per ml.

Preparation of integrated substrate DNA. The C3H mammary carcinoma cell line, 34I cl 101, was utilized for preparation of integrated MMTV substrate DNA. High-molecular-weight DNA (>50,000 base pairs) was prepared from confluent monolayers by the procedure of Gross-Bellard et al. (11).

Preparation of unintegrated substrate DNA. Closed circular MMTV DNA molecules were obtained from M714H cl 3, a cloned HTC line which had been chronically infected at a high multiplicity of infection with MMTV. Low-molecular-weight DNA was extracted by the Hirt procedure (17). Form I DNA was purified on two successive propidium iodide-cesium chloride gradients. Gradient fractions were analyzed by restricting 10% of the fraction with *Eco*RI, electrophoresing the fractions on a 0.7% agarose gel, blotting on nitrocellulose, and hybridizing the Southern transfer to MMTV representative probe $(cDNA)_{rep}$. The appropriate fractions were pooled, butanol ex-

tracted, dialyzed, phenol extracted, and ethanol precipitated. Further restriction mapping confirmed the identity of the circular DNA with the map published by Shank et al. (41).

Restriction endonuclease digestions. Restriction endonucleases were purchased from Bethesda Research Laboratories and New England Biolabs. Reaction conditions were those recommended by the manufacturers, except that the DNA was prepared by incubation of 20-μg samples of DNA with 2 U of *Eco*RI for various lengths of time, from 5 min to 5 h. Agarose gel electrophoresis of 1-μg samples of each partially restricted DNA was used to determine conditions optimal for generating 15- to 25-kilobase DNA fragments.

Gel electrophoresis and hybridization analysis. Analytical electrophoresis of restricted DNA samples was carried out in a horizontal gel apparatus with 0.4 or 1.0% agarose gels containing 20 mM sodium acetate, 40 mM Tris-hydrochloride (pH 8.0), 2 mM EDTA, and 0.5 μg of ethidium bromide per ml. DNA to be hybridized with [^{32}P]cDNA was transferred from the gel to nitrocellulose filters by the procedure of Southern (45) with minor modifications. MMTV [^{32}P]cDNA$_{rep}$ was synthesized in a reaction catalyzed by avian myeloblastosis virus reverse transcriptase and primed with oligomers of calf thymus DNA (47). For hybridization of DNA on nitrocellulose filters the procedure of Jeffreys and Flavell (20) was utilized.

Preparation of phage DNA for cloning. Bacteriophage vectors Charon 4A and λgtWES·λB were propagated, and DNA was purified as previously described (13). Preparation of vector arms was as described previously (13, 28), except that λgtWES·λB was cleaved with *Sac*I (New England Biolabs) rather than with *Eco*RI.

In vitro packaging and detection of phage recombinants. Preparation of packaging extracts and in vitro packaging protocols have been previously described (3, 13). MMTV-positive phage recombinants were detected, subcloned, and analyzed as previously described (2, 13).

Biological and physical containment. The EK2-certified vectors λgtWES·λB and Charon 4A were utilized in conjunction with in vitro packaging under the conditions prescribed in the National Institutes of Health Guidelines for Recombinant DNA Research. All packaging reactions, plating, and propagation of recombinant phages were carried out in certified P2 facilities at the National Institutes of Health. Plasmid transformations, colony assays, and plasmid amplification steps were also carried out in P2 facilities. *E. coli* host strains LE392, DP50supF, and χ1776 used in these experiments were all EK2 certified.

Bacterial transformations. Plasmid pBR322 was used in the plasmid subcloning experiments. Restriction fragments from MMTV-containing recombinant phages were ligated with *Bam*HI-, *Eco*RI-, and *Sac*I-restricted plasmids under the ligation conditions previously described (13).

Transformation of *E. coli* χ1776 with recombinant plasmids was as described by Hanahan (personal communication). Briefly, χ1776 was grown in X broth to mid-log phase (optical density at 550 nm, 0.2). Cells were centrifuged for 12 min at 2,000 rpm, drained, and suspended in 0.33 volume of transformation buffer (35 mM KAc [pH 5.8], 100 mM RbCl, 45 mM $MnCl_2$, 10 mM $CaCl_2$, 0.5 mM LiCl, 15% sucrose) and left on ice for 5 min. The suspension was centrifuged for 8 min at 2,000 rpm (4°C), suspended in 0.04 volume of transformation buffer, divided into samples in tubes (12 by 75 mm), and incubated on ice for 15 to 20 min before adding 7 μl of dimethyl sulfoxide. The samples were swirled gently on ice for another 20 min before addition of 1 to 10 μl of the ligated plasmid DNA (ca. 250 ng of DNA). The bacteria were swirled on ice for 20 min, placed in a −50°C bath for 100 s, thawed, and placed on ice for 10 min. The tubes were then incubated at 40 to 42°C for 60 s before adding 400 μl of X broth.

Strain N99 was transformed according to the procedure of Mandel and Higa (27).

Detection of plasmid recombinants. Amp^r Tet^s colonies were assayed for MMTV-containing plasmids by the colony hybridization technique of Grunstein and Hogness (12).

Amplification and purification of plasmid DNA. Bacterial colonies containing MMTV plasmid recombinants were amplified by the following modified procedure of Boyer (personal communication). MMTV-positive colonies were inoculated in 1 liter of medium and grown at 37°C to an optical density at 550 nm of 0.5 to 0.7. The bacteria were harvested by low-speed centrifugation at 4°C. The bacterial pellets were suspended in 25% sucrose–0.5 M Tris-hydrochloride (pH 8)–20 mM ETA. Four milliliters of lysozyme solution (10 mg of lysozyme per ml, 0.25 M Tris-hydrochloride [pH 8], 20 mM EDTA) was added, and the extract was incubated at 4°C for 5 min; 8 ml of 0.25 M EDTA was added, and the extract was swirled at 4°C for 5 min; 32 ml of 0.4% Triton X-100–0.5 M Tris-hydrochloride (pH 8)–0.05 M EDTA was added, followed by incubation at 4°C for 10 min. The lysates were centrifuged at 25,000 rpm for 1 h at 4°C, and the supernatants were collected and extracted with 1 volume of phenol saturated with 0.1 M Tris-hydrochloride (pH 9). The aqueous phase was then ethanol precipitated and dissolved in 3 ml of 0.01 M Tris-hydrochloride (pH 7.8)–0.01 M EDTA–0.2 M NaCl before adding RNase A to 50 μg/ml and incubating at 37°C for 15 min. The sample was then applied to a 30-ml agarose A50m column equilibrated with sample buffer. Fractions were monitored by optical density at 260 nm; plasmid DNA chromatographed in the void volume. The plasmid fractions were pooled, ethanol precipitated, and dissolved in 0.01 M Tris-hydrochloride (pH 7.8)–0.01 M EDTA. CsCl was added to 50% (wt/wt), and propidium iodide was added to 100 μg/ml. The gradient was centrifuged at 45,000 rpm for 24 to 36 h at 20°C in a VTi50 rotor. DNA bands corresponding to plasmid forms I and II were collected, pooled, and extracted four times with 50% CsCl-saturated isopropanol. The DNA was then dialyzed against 0.01 M Tris-hydrochloride (pH 7.4)–1 mM EDTA.

Preparation of small DNA fragments for sequencing. Restriction fragments for sequencing were purified by a modification of the procedure of Maxam and Gilbert (30, 31). Plasmid DNA (400 to 800 μg) was restricted with 800 U of an appropriate restriction enzyme. The resulting fragments were separated by preparative electrophoresis on vertical 4.0% polyacrylamide slab gels (3 mm thick; Hoefer electrophoresis

apparatus) for 5 h at 50 V in 100 mM Tris-borate (pH 8.3)-1 mM EDTA. After electrophoresis, the slab gels were stained for 10 min in 5 μg of ethidium bromide per ml, and the DNA bands were visualized with a UV light source. Fragments of interest were sliced out of the gel and added to 10 to 25 ml of 0.01 M Tris-hydrochloride (pH 7.4)-1 mM EDTA. After mincing of the polyacrylamide, DNA was eluted at 25°C by gentle mixing overnight. The solution was centrifuged at low speed, and the DNA-containing supernatant was loaded onto a small DE52 anion exchange column (Whatman) equilibrated with 0.01 M Tris-hydrochloride (pH 7.4)-1 mM EDTA. DNA was eluted from the column with 1 ml of 1 M NaCl, ethanol precipitated, and dissolved in 200 μl of 10 mM Tris-hydrochloride (pH 7.4)-1 mM EDTA.

DNA sequencing. DNA sequencing was performed as described by Maxam and Gilbert (30, 31) without modification. DNA fragments were 5′ end labeled by using previously described protocols (31). The sequencing reactions used were the G, G+A, A>G, C+T, C, and A>C reactions. Thin gels (0.4 mm) (40) 80 and 40 cm in length were used in the gel electrophoresis steps.

RESULTS

Molecular cloning of integrated MMTV. The substrate DNA for molecular cloning of integrated MMTV was obtained from an MMTV-producing cloned cell line (34I cl 101) previously derived from a mammary carcinoma of a C3H/Crgl mouse (33, 34). This cell line contains a complement of 20 to 25 copies of exogenously introduced milk-born MMTV in addition to the endogenous sequences present in the C3H genome (7, 34). Since the strain of MMTV horizontally transmitted in C3H mice contains a single internal *Eco*RI site (41), our cloning strategy involved partial *Eco*RI restriction of genomic DNA, with the expectation that some of the MMTV recombinants would contain a complete genome along with adjacent cellular sequences. *Eco*RI fragments with a size distribution of 10 to 20 kilobases were generated and inserted into the Charon 4A lambda phage vector, and MMTV-specific recombinants were identified (76).

Although it was anticipated that some of the recombinants would contain complete integrated MMTV genomes, only clones containing the 3′ half of the MMTV genome were actually identified. Numerous attempts to clone an accurate copy of the milk-born or "exogenous" MMTV genome have failed on either integrated or unintegrated substrates. We believe that the left half of the MMTV genome contains a DNA sequence that cannot be readily cloned in *E. coli* vector systems. Other investigators have reported similar observations (Majors and Varmus, 9th Annual UCLA-ICN Symposium on Animal Virus Genetics, in press).

In two independent cloning experiments, a total of 86 MMTV recombinants were identified; restriction maps for 18 of these have been determined. Fifteen of these recombinants contained 3′ halves corresponding to the exogenous MMTV genome. The other three recombinants represent copies of MMTV endogenous to the C3H genome as determined by comparison with endogenous restriction maps (7; L. D. Johnson et al., manuscript in preparation; L. Donehower, unpublished observations) and will not be discussed here. Restriction maps of three of the exogenous MMTV 3′ clones are compared in Fig. 2 to the map of the unintegrated linear form of MMTV (41). Figure 2 indicates the virtual identity of the integrated 3′ halves with the 3′ end of the unintegrated linear form of MMTV; the structure of the MMTV integrated sequences is colinear and nonpermuted with respect to the unintegrated form. These mapping data, together with the sequencing information presented below, confirm the exogenous origin of these provirus inserts. Additionally, the adjacent nonviral cellular sequences are unique for each independent clone, in agreement with earlier observations by Southern blot analyses that the site of integration is random with respect to cell DNA but specific with respect to the viral genome.

Molecular cloning of the MMTV LTR from unintegrated circles. To prepare an unintegrated circular DNA substrate for cloning, we took advantage of the observation of Ringold et al. (36) that nonmurine cell lines infected with MMTV continue to replicate DNA intermediates during chronic growth of the cells. Variants of the HTC cell line developed by Thompson et al. (48) were infected with MMTV (C3H); cell clones actively producing MMTV RNA were identified. A preliminary characterization of these clones has been reported (49). Form I circular DNA was purified as described above. To avoid difficulties associated with cloning the 5′ half of MMTV in *E. coli* systems (see discussion above), we chose to cleave permuted copies of the MMTV LTR from circles containing two or more LTR sequences by restriction of the circular DNA with *Sac*I, an enzyme that cuts once within the MMTV LTR (41), and not in the unique sequences (Fig. 3). The resulting fragments were ligated into λWES·λB cut previously with *Sac*I. A clone containing a 1.3-kilobase fragment that hybridized with MMTV $cDNA_{rep}$ was identified; the restriction map was identical to that for the terminal redundancy of the integrated MMTV recombinants permuted at the *Sac*I site.

Subcloning of integrated and unintegrated MMTV in plasmids. *Bam*HI-*Eco*RI

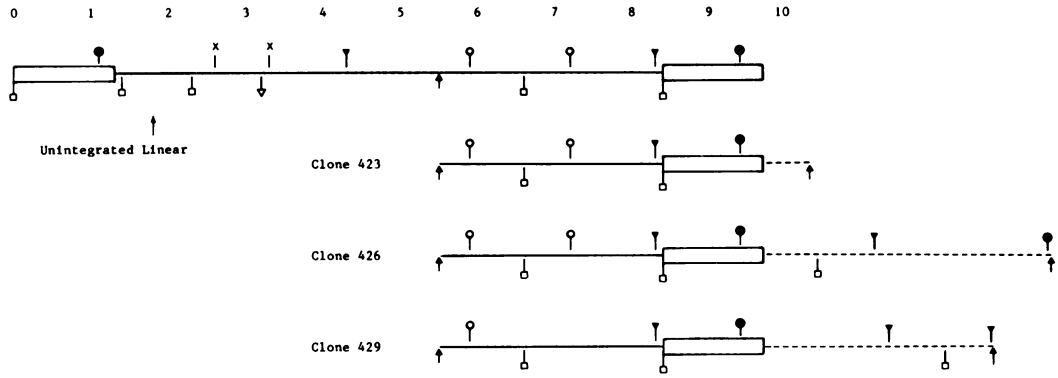


FIG. 2. *Restriction maps of integrated MMTV clones. The restriction maps for three Charon 4A recombinants containing the 3′ half of MMTV (C3H) and adjacent cellular DNA are shown. The clone maps are compared with the restriction map of the unintegrated linear form of MMTV (41) shown at the top. Solid lines represent DNA which hybridizes to MMTV probe; boxed regions delimit the LTR; broken lines correspond to nonhybridizing mouse cellular sequences. Symbols for restriction endonuclease sites are as follows: □, PstI; ●, SacI; ×, XhoI; ▽, KpnI; ▼, BglII; ↑, EcoRI; ○, BamHI.*

fragments spanning the MMTV rightward LTR and the virus-host joint from three integrated MMTV 3′ bacteriophage clones were subcloned in pBR322. Transformants of *E. coli* host strain χ1776 were selected by screening for tetracycline sensitivity and colony hybridization against an MMTV [^{32}P]cDNA$_{rep}$ probe, followed by restriction mapping of DNA from the MMTV-positive colonies.

The bacteriophage clone containing the permuted MMTV unintegrated LTR was restricted with *Sac*I, and the fragments were ligated into pBR322 containing a b2 DNA fragment of bacteriophage λ inserted at the *Eco*RI and *Hin*dIII sites. This molecule (kindly provided by A. Das) contains a single *Sac*I site into which the unintegrated *Sac*I redundancy fragment was inserted.

Sequencing strategies. DNA fragments from the three integrated and one unintegrated MMTV-containing plasmids were sequenced by the strategy outlined in Fig. 4. An example of the sequencing data for two virus-host joints is given in Fig. 5. In the case of 4a1-1 (Fig. 4), all regions of the terminal redundancy have been sequenced at least twice from overlapping fragments. The only site not covered by overlapping sequences is the *Sac*I site at position −240. It is therefore possible that two very closely spaced *Sac*I sites occur in this region. Although this unlikely possibility will be checked by further sequence analysis, neither the putative transcriptional regulatory sequences nor the potential protein-encoding sequence is included in this region of the LTR, and this minor point is not critical to the central arguments of this paper.

Occasional ambiguities in the sequence were eliminated by the use of five sequencing reactions (an additional A>C) per DNA fragment instead of the customary four. The redundancies of the other two integrated clones and the unintegrated clone provided confirmation of the LTR sequence. We emphasize that all four clones were identical in sequence in all of the MMTV-containing regions examined. This perfect congruity supports the conclusion that each of the three integrated clones represents an exogenous genome rather than an endogenous provirus.

MMTV terminal redundancy sequences. The integrated MMTV LTR sequence for clone 4a1-1 is shown in Fig. 6a; as discussed above, LTR sequences for the other integrated clones were identical over the regions that were compared. Adjacent cellular sequences for the three integrated clones are given in Fig. 6b. The DNA sequence across the joint between the two copies of the LTR from the unintegrated circle clone is compared with the integrated LTR in Fig. 3.

DISCUSSION

Size and boundaries of the MMTV terminal redundancy. Three criteria have been employed to delineate the boundaries of MMTV terminal redundancy. Initially, the rightward end was defined by the point of sequence divergence among the three integrated clones (Fig. 6b). The second approach involves the comparison of the integrated and unintegrated redundancy sequences as shown in Fig. 3. The joint region of the unintegrated circle is thought to involve the covalent joining of the 3′ and 5′ ends of the viral linear DNA, thereby representing

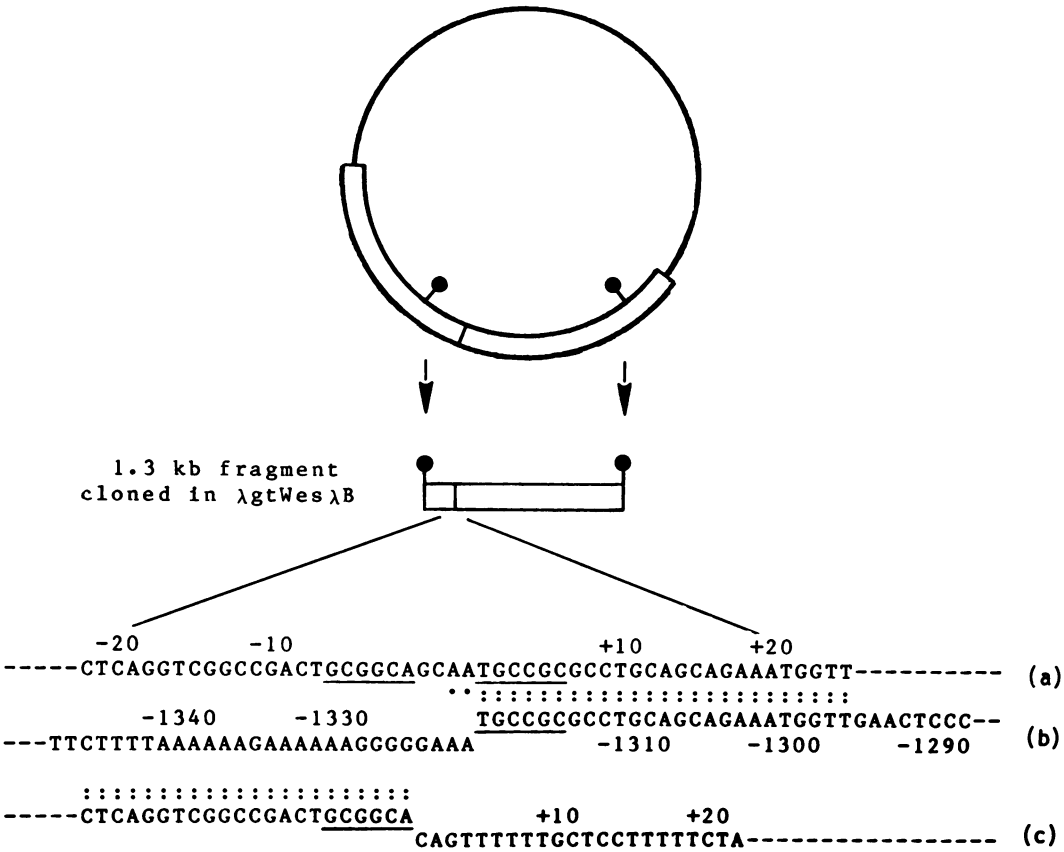


FIG. 3. *Cloning of unintegrated circular forms of MMTV. This diagram shows the strategy used in cloning the LTR region of the MMTV circles. Purified circles were cleaved with SacI which resulted in fragments of 1.3 and 7.7 kilobases (kb). The 1.3-kb fragment contained the joint of the 5′ and 3′ ends of the viral DNA in the linear form and one entire copy of the MMTV LTR in a permuted form. The fragments were ligated with λgtWES·λB arms previously restricted with SacI, and recombinants were generated. A recombinant phage containing the 1.3-kb fragment was isolated, the DNA was restricted with SacI, and the 1.3-kb fragment was further subcloned in a modified form of pBR322. The DNA sequence (a) for the 3′-5′ joint region in this fragment is compared with the left end (b), and right end (c), of the 3′ LTR from the integrated provirus (clone 4a1-1 of Fig. 6).*

the joining of the 5′ end of one terminal redundancy to the 3′ end of the other terminal redundancy. Comparison of the unintegrated joint to sequences at the left and right ends of the integrated terminal redundancy should therefore delimit the boundaries of the integrated LTR. When this comparison is made (Fig. 3), the sequences of the integrated LTR left and right ends are found to agree with the sequences present at the 3′-5′ circle joint, with the exception of two to four bases present in the circle joint, but missing at the integrated LTR ends. The ambiguity arises because the GC pair of residues is clearly absent when all three virus-host joints are compared (see Fig. 6b), whereas the AA pair may be present at the left end of the integrated LTR.

Majors and Varmus (9th Annual UCLA-ICN Symposium on Animal Virus Genetics, in press) have determined that the AA pair potentially defining the LTR left end is absent at the leftward host-virus joint of a partially deleted MMTV provirus. The exact composition of the rightward host-virus joint for this partial provirus is ambiguous because a G residue at the joint could derive either from the viral DNA or from the duplicated cell sequence. Comparison of their rightward joint with the three host-virus joints presented here clearly indicates, however, that the G must have originated from the transposed cell sequence. It seems clear, therefore, that the four bases present at the unintegrated circle joint represent the original composition of the linear precursor, and that two bases are

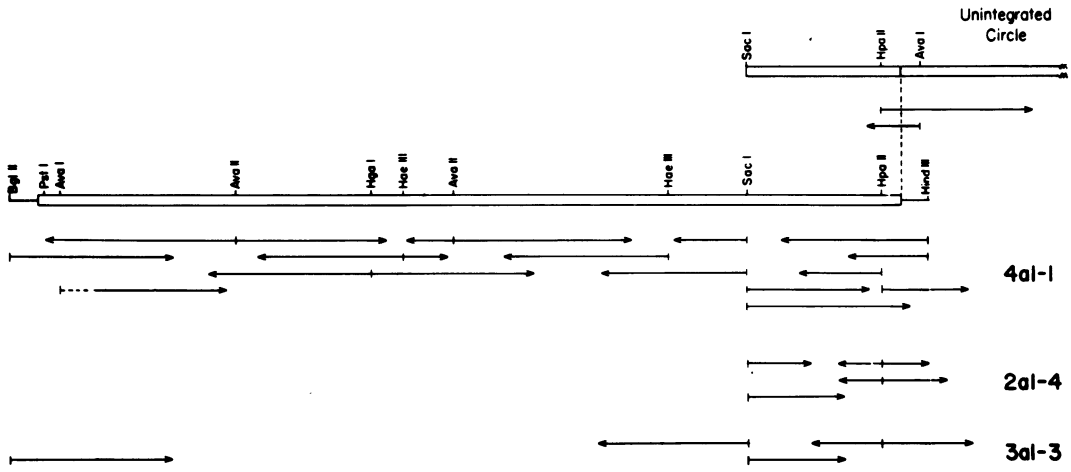


FIG. 4. *Strategies used in determining MMTV DNA sequences. This figure indicates regions of the unintegrated and integrated LTR (along with adjacent cellular sequences) which were sequenced by the method of Maxam and Gilbert. As in Fig. 1, the boxed region represents the boundaries of the LTR. Horizontal arrows indicate the direction and extent of sequence determination. The clone designations correspond to BamHI-EcoRI subclones in pBR322 of the original EcoRI inserts (Fig. 2); 2a1-4 is from clone 423, 3a1-3 is from clone 426, and 4a1-1 is from clone 429.*

deleted from the left end of the 5′ LTR and two bases from the right end of the 3′ LTR during the integration event. Integration of retroviral DNA, therefore, is precise to the single nucleotide with respect to the viral genome. The same conclusion has been reached by Shimotohne et al. (43) for the integration of spleen necrosis virus. The resultant size for the integrated MMTV terminal redundancy is 1,325 base pairs, and the size of the unintegrated counterpart is 1,327 base pairs.

A third feature defining the boundaries of the LTR is the presence of 6-base-pair complementary palindromes (Fig. 3). Similar inverted repeat sequences, varying from 3 to 11 nucleotides, have recently been identified in the LTR sequences of Moloney sarcoma virus (7a), Moloney murine leukemia virus (44), avian sarcoma virus (R. Swanstrom et al., 9th Annual UCLA-ICN Symposium on Animal Virus Genetics, in press), spleen necrosis virus (43), and independently reported for MMTV by Majors and Varmus (in press). This feature, together with the duplication of cellular sequences at both ends of integrated retroviruses, presents a remarkable structural analogy with that of procaryotic transposition elements and forms the basis of current postulates regarding the mechanistic similarity of retroviral integration and procaryotic transposition (5, 10).

Potential transcription regulatory sites in the MMTV LTR. (i) Initiation signals. Comparison of DNA sequences proximal to the mRNA cap addition site for a number of eucaryotic genes reveals the frequent occurrence of a common sequence, TATAAAA or a close variant, 25 to 30 nucleotides upstream from the RNA initiation point (9, 21, 23, 32, 50, 54). This sequence is hypothesized to be an RNA polymerase II initiation signal, analogous to the so-called "Pribnow box" of procaryotic promotors (35). Although other examples of eucaryotic genes whose expression is mediated by RNA polymerase II have been described that lack the upstream TATAAAA box analog (1), it has been elegantly demonstrated that initiation in vitro in cell-free transcription systems can occur at sites with a proximal TATAAAA sequence (29, 52).

The length of the 5′ strong stop sequence for MMTV has been independently estimated to be 140 to 150 nucleotides (15; J. Taylor and H. Young, personal communication). Referring to the topology for integrated viral DNA summarized in Fig. 1, it would be predicted that the initiation site for MMTV transcription should occur between positions −140 and −150, as shown in Fig. 6a. An RNA polymerase II initiation signal might therefore be expected between positions −165 and −180. A heptanucleotide, TATAAAA, is in fact located at position −167 and would therefore fit the pattern described for many RNA polymerase II-mediated genes. Another remarkable group of nucleotides occurs between positions −195 and −174. The 10-base-pair cluster (CTTATGTAAA), similar to the TATAAAA sequence, is repeated precisely. Determination of whether these sequences have any functional significance awaits the demon-

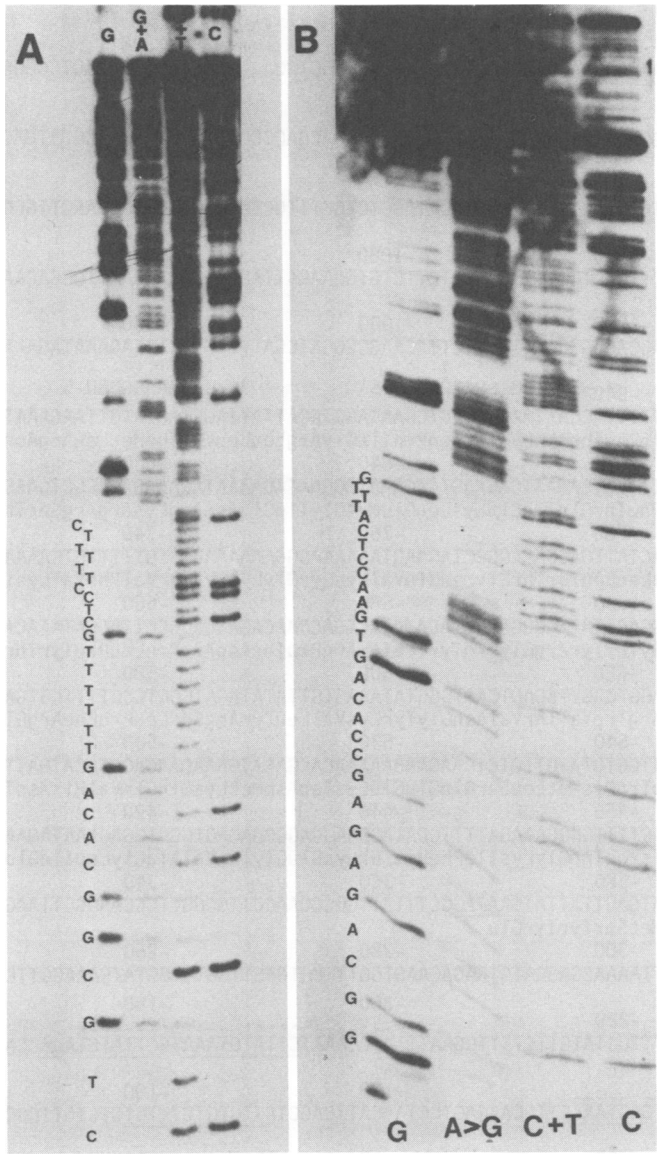
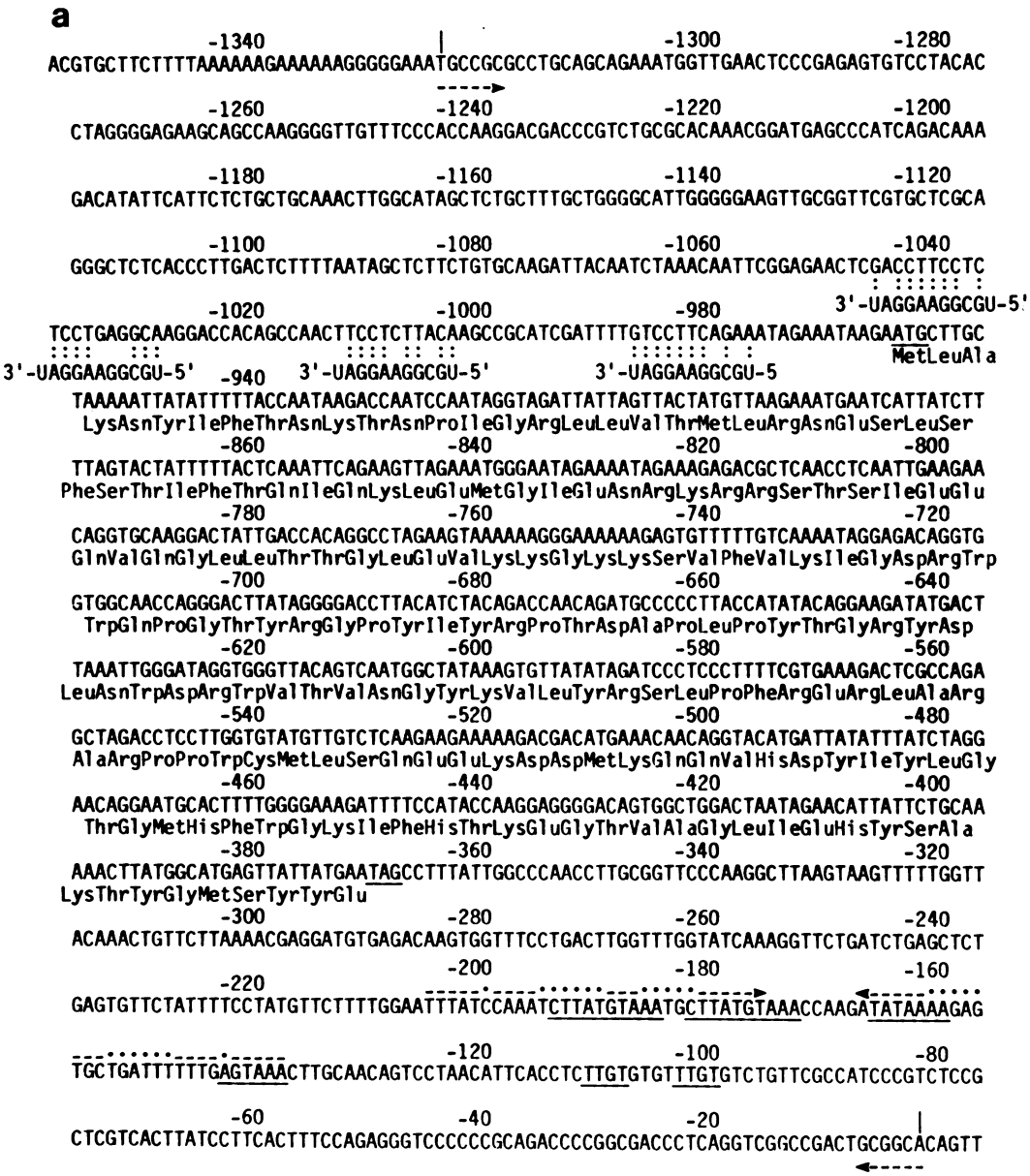


FIG. 5. *Autoradiogram of sequencing gels illustrating sequence divergence at the virus-host joints in independent MMTV integration events. (a) A 1,900-base-pair (bp) fragment isolated from plasmid 4a1-1 containing the 3' integrated virus-host joint was restricted with HpaII, 5' end labeled, secondarily restricted with AvaI, and separated on an 8% polyacrylamide gel. A 350-bp fragment containing the HpaII end-labeled joint was purified from the gel and sequenced. (b) A 1,700-bp fragment purified from plasmid 2a1-4 containing the 3' virus-host joint was restricted with HpaII, 5' end labeled, restricted with AvaI, and electrophoresed on a 4% polyacrylamide gel. A 1,000-bp HpaII end-labeled fragment was purified and sequenced.*

stration of specific viral RNA initiation with these or similar DNA templates. Other work in this laboratory (L. D. Johnson et al., in preparation) has indicated that the 5' ends of viral RNAs map within 2 nucleotides of position −136, 21 to 25 nucleotides downstream from the Hogness box.

(ii) Termination signals. Examination of DNA sequences available to date in regions of eucaryotic DNA where RNA termination and polyadenylation occur reveals the presence of three common features.

(a) Sequences with hyphenated dyad symmetry are commonly found proximal to the point


a
-1340
-1300
-1280
ACGTGCTTCTTTTAAAAAAGAAAAAAGGGGGAAATGCCGCGCCTGCAGCAGAAATGGTTGAACTCCCGAGAGTGTCCTACAC
-1260
-1240
-1220
-1200
CTAGGGGAGAAGCAGCCAAGGGGTTGTTTCCCACCAAGGACGACCCGTCTGCGCACAAACGGATGAGCCCATCAGACAAA
-1180
-1160
-1140
-1120
GACATATTCATTCTCTGCTGCAAACTTGGCATAGCTCTGCTTTGCTGGGGCATTGGGGGAAGTTGCGGTTCGTGCTCGCA
-1100
-1080
-1060
-1040
GGGCTCTCACCCTTGACTCTTTTAATAGCTCTTCTGTGCAAGATTACAATCTAAACAATTCGGAGAACTCGACCTTCCTC
-1020
-1000
-980
3'-UAGGAAGGCGU-5'
TCCTGAGGCAAGGACCACAGCCAACTTCCTCTTACAAGCCGCATCGATTTTGTCCTTCAGAAATAGAAATAAGAATGCTTGC
MetLeuAla
3'-UAGGAAGGCGU-5'
-940
3'-UAGGAAGGCGU-5'
3'-UAGGAAGGCGU-5
TAAAAATTATATTTTTACCAATAAGACCAATCCAATAGGTAGATTATTAGTTACTATGTTAAGAAATGAATCATTATCTT
LysAsnTyrIlePheThrAsnLysThrAsnProIleGlyArgLeuLeuValThrMetLeuArgAsnGluSerLeuSer
-860
-840
-820
-800
TTAGTACTATTTTTACTCAAATTCAGAAGTTAGAAATGGGAATAGAAAATAGAAAGAGACGCTCAACCTCAATTGAAGAA
PheSerThrIlePheThrGlnIleGlnLysLeuGluMetGlyIleGluAsnArgLysArgArgSerThrSerIleGluGlu
-780
-760
-740
-720
CAGGTGCAAGGACTATTGACCACAGGCCTAGAAGTAAAAAAGGGAAAAAAGAGTGTTTTTGTCAAAATAGGAGACAGGTG
GlnValGlnGlyLeuLeuThrThrGlyLeuGluValLysLysGlyLysLysSerValPheValLysIleGlyAspArgTrp
-700
-680
-660
-640
GTGGCAACCAGGGACTTATAGGGGACCTTACATCTACAGACCAACAGATGCCCCCTTACCATATACAGGAAGATATGACT
TrpGlnProGlyThrTyrArgGlyProTyrIleTyrArgProThrAspAlaProLeuProTyrThrGlyArgTyrAsp
-620
-600
-580
-560
TAAATTGGGATAGGTGGGTTACAGTCAATGGCTATAAAGTGTTATATAGATCCCTCCCTTTTCGTGAAAGACTCGCCAGA
LeuAsnTrpAspArgTrpValThrValAsnGlyTyrLysValLeuTyrArgSerLeuProPheArgGluArgLeuAlaArg
-540
-520
-500
-480
GCTAGACCTCCTTGGTGTATGTTGTCTCAAGAAGAAAAAAGACGACATGAAACAACAGGTACATGATTATATTTATCTAGG
AlaArgProProTrpCysMetLeuSerGlnGluGluLysAspAspMetLysGlnGlnValHisAspTyrIleTyrLeuGly
-460
-440
-420
-400
AACAGGAATGCACTTTTGGGGAAAGATTTTCCATACCAAGGAGGGGACAGTGGCTGGACTAATAGAACATTATTCTGCAA
ThrGlyMetHisPheTrpGlyLysIlePheHisThrLysGluGlyThrValAlaGlyLeuIleGluHisTyrSerAla
-380
-360
-340
-320
AAACTTATGGCATGAGTTATTATGAATAGCCTTTATTGGCCCAACCTTGCGGTTCCCAAGGCTTAAGTAAGTTTTTGGTT
LysThrTyrGlyMetSerTyrTyrGlu
-300
-280
-260
-240
ACAAACTGTTCTTAAAACGAGGATGTGAGACAAGTGGTTTCCTGACTTGGTTTGGTATCAAAGGTTCTGATCTGAGCTCT
-200
-180
-160
-220
GAGTGTTCTATTTTCCTATGTTCTTTTGGAATTTATCCAAATCTTATGTAAATGCTTATGTAAACCAAGATATAAAAGAG
-120
-100
-80
TGCTGATTTTTTGAGTAAACTTGCAACAGTCCTAACATTCACCTCTTGTGTGTTTGTCTGTTCGCCATCCCGTCTCCG
-60
-40
-20
CTCGTCACTTATCCTTCACTTTCCAGAGGGTCCCCCCGCAGACCCCGGCGACCCTCAGGTCGGCCGACTGCGGCACAGTT
b
-20
-10
-1
+10
+20
+30
4a1-1
--CAGGTCGGCCGACTGCGGCACAGTTTTTTGCTCCTTTTTCTAGATGTAAT
3a1-3
--CAGGTCGGCCGACTGCGGCAGCAGAAGATAATGCTCTCACAGCACAATTC
2a1-4
--CAGGTCGGCCGACTGCGGCAGAGAGCCACAGTGAACCTCATTTGAACACC
+30
+40
+50
+60
+70
+80
4a1-1
TTTTAAAGCTTATTTTTTAACTTTCACATGTGCTACACTCACATGTGCAATGAGTGAACAG--
3a1-3
TTTTCCTAACCAAATGCGGCAAATCACTGTGCCAAATGCGGCAAATCACTGTGTGGTTCCT--
2a1-4
ACAGGGCTTTTTGACACATGAGCTAATGGATTGCATGCATAAGCCCTACATA----------

at which the RNA is terminated. Similar structures are found proximal to termination points for procaryotic mRNA's (38) and are thought to result in the formation of "stem-loop" structures in the nascent RNA transcript that may be recognized in the termination event. Indeed, sequence variations in mutants with altered termination phenotypes support this model (46). Although the 3′ end of MMTV genomic RNA has not been described, preliminary evidence (Diggelmann et al., personal communication) indicates that the molecule is terminated approximately at position −120 in Fig. 6a. A large complementary palindrome can be constructed in the sequence proximal to this termination point (shown in Fig. 6a as the broken overlining). Although the participation of this structure in termination is purely speculative, its involvement would suggest a possible resolution of the 5′ termination problem discussed earlier. Initiation of RNA synthesis would occur in the leftward LTR approximately at position −140; the left half of the dyad symmetry would therefore not be present at the 5′ end of viral RNA, and termination would not occur. Such an argument has been presented for the termination region of Moloney murine leukemia virus by Dina and his colleagues (personal communication).

(b) A second feature common to many termination sites is the presence of a tetranucleotide, TTTT or TTGT (often repeated), distal to the point of polyadenylic acid addition (21–23). Again referring to the sequence in Fig. 6a, there are two TTGT tetranucleotides located downstream from the putative termination site at positions −110 and −102. Although these clusters have not been functionally implicated in the termination of polymerase II-directed transcripts, Brown and his colleagues (4) reported that the mutation of one base in a nonrepeated TTTT cluster downstream from a polymerase III-directed 5S transcript led to the loss of termination.

(c) Notably absent in the region of the LTR sequence where viral RNA is terminated is the canonical polyadenylation signal AATAAA. Not only is this hexanucleotide frequently found 11 to 19 nucleotides proximal to the polyadenylic acid addition site in mRNA's, but also an elegant analysis of site-directed deletion mutants at the polyadenylation site for late simian virus 40 message by Fitzgerald and Shenk (submitted for publication) provides strong experimental evidence for the functional role of the signal. It should be noted in this regard that the AATAAA signal does occur in the predicted position in the LTR for the type C retroviruses, Moloney murine sarcoma virus (7a) and avian sarcoma virus (53). A close analog, AGTAAA, exists in the MMTV sequence at the appropriate position (−132). This group would form the base of the stem-loop described above.

Protein-coding potential. Approximately 1,200 nucleotides of the MMTV LTR are derived from information uniquely present at the 3′ end of viral RNA. It was of interest to determine whether a sequence of this size (representing 14% of the nonredundant information of the genome) contained potential protein-encoding information. We therefore analyzed the LTR sequence for the existence of a translational open reading frame. Figure 7 shows that there is, in fact, a large open reading frame extending from nucleotides −1,161 to −370, a total of 792 nucleotides. The first initiation codon available in this open reading frame occurs at position −963, resulting in a sequence of 594 nucleotides coding for a hypothetical polypeptide of 198 amino acids. The nucleotide sequence of this open reading frame and the corresponding amino acid sequence is shown in Fig. 6a.

The probability of an open reading frame of this size occurring by random chance is quite low, less than 10^{-5}. As mentioned above, extensive redundant sequence comparisons were performed to confirm the sequence as presented. Sequencing errors, furthermore, are much more likely to result in loss of a translational frame rather than creation of a nonexistent open frame.

It is reasonable to assume that a message for the putative protein would be transcribed or processed separately from the *gag-pol* and *env*

FIG. 6. *Nucleotide sequences of three integrated and one unintegrated MMTV clone. (a) The nucleotide sequence of the 4a1-1 MMTV LTR along with adjacent unique MMTV and cellular sequences is shown. The sequence is numbered in a negative direction from the putative virus-host joint. The 6-base complementary palindromes at the ends of the LTR are denoted by dashed arrows. Proposed transcriptional and translational regulatory signals are underlined or overlined and referred to in the text. Amino acids encoded by a large open reading frame in the LTR are positioned below the relevant codons. Regions of the MMTV LTR sequenced in 3a1-3, 2a1-4, and the unintegrated circle are identical in sequence to the 4a1-1 LTR and have been omitted here. (b) The nucleotide sequences of three 3′ MMTV virus-host joints are compared. Numbering begins at the putative joint in a negative direction into MMTV and in a positive direction into the cellular sequences. The perfect homology among the MMTV-containing sequences is denoted by double dots. Tandem 22-base-pair (bp) direct repeats in clone 3a1-3 and 10-bp repeats in 4a1-1 are overlined. The 7-bp homologies to the right end of the LTR are underlined.*

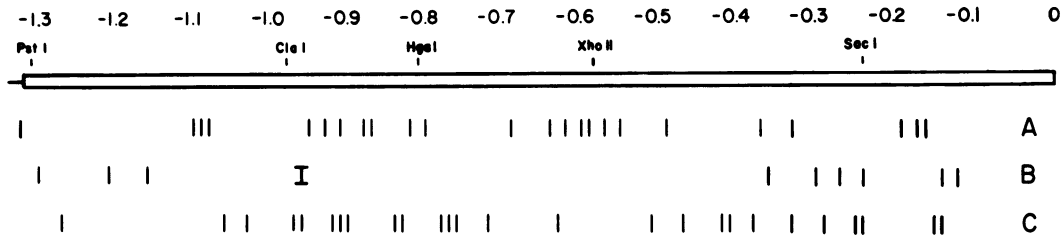


FIG. 7. *Reading frame analysis of the MMTV LTR. An abbreviated restriction map of the MMTV LTR is shown with the standard numbering system of Fig. 6a. Below the LTR map, vertical lines represent the position of translation termination codons (TAA, TGA, TAG) found in the three different reading frames designated A, B, and C. The first initiation codon (ATG) available in the large open reading frame (B phasing) is indicated by an I.*

gene complexes during formation of the various MMTV mRNA's. Under this scenario, we would expect to find either RNA polymerase II initiation signals or an acceptor splice junction proximal to the coding sequence. We find no obvious upstream promotor analogs, but this absence does not eliminate the possibility of independent promotion, as discussed above. Sequences can be identified proximal to the open reading frame (for example, at nucleotides −1,112 to −1,124) that are analogous to an acceptor splice junction (26), but variation in the acceptor consensus sequence is too great to assign statistical significance to these homologies.

We have also analyzed the DNA sequence upstream from the putative initiator codon for potential ribosome binding sites. Hagerbuchle and collaborators (14) have observed a sequence at the 3′ end of the 18S rRNA subunit that is conserved in eucaryotes. Homology between this sequence and nontranslated leader regions of several eucaryotic messages was noted. These investigators postulate the potential involvement of this conserved sequence in message binding, analogous to the similar "Shine-Dalgarno" sequence in procaryotic systems. Fig. 6a shows that four sites exhibiting strong homology with the conserved 3′ 18S rRNA sequence of Hagerbuchle et al. are within 100 nucleotides upstream of the initiator codon. Computer analysis of the complete sequence reveals that homology with this cluster is highly nonrandom in the LTR.

A biological role for this putative protein in MMTV-infected cells can only be the subject of speculation at this time. It has been hypothesized that this region of the viral genome encodes the function responsible for mammary carcinogenesis (16), but the evidence for this postulate is limited. Alternatively, one could propose that the polypeptide is involved in hormone induction of transcription, a property unique for MMTV among retroviruses.

The known polypeptides encoded by the MMTV genome do not include an obvious candidate for this sequence. Peters and Dickson (9th Annual UCLA-ICN Symposium on Animal Virus Genetics, in press) have reported that in vitro translation of the 3′ end of MMTV genomic RNA generates a 30,000- to 40,000-molecular-weight protein. This would be too large for the reading frame described, unless it represents one exon of a spliced message. An obvious approach to resolving this issue is to synthesize part or all of the predicted polypeptide and determine its presence in infected cells. Such efforts are under way.

Integration sites. Restriction endonuclease analysis in a variety of systems (6, 19, 37, 39) has demonstrated that exogenously introduced retroviral genomes integrate at a wide variety of sites in the host cell genome. This evidence indicates that either a cellular sequence-specific integration mechanism is not operative or the recognition sequence is too small to be detected by restriction endonuclease analysis. The recently discovered analogy between the structure of integrated proviruses and procaryotic transposition elements also argues that direct sequence homology between retroviral and host DNA does not play a role in the integration process. Transposition elements can integrate at a large number of sites in bacterial DNA; the utilization of a specific site with perfect sequence homology between the host site and the integrating molecule (as in lambda bacteriophage integration [24]) clearly does not occur. Recent evidence obtained by sequence analysis of a large number of transposition events for the Tn*3* insertion element, however, suggests that homology between the ends of the transposition element and the recipient DNA may be in some way involved in the integration process (51). We have therefore examined the DNA sequences obtained for the 3′ virus-cell joints of three independent integration events for potential ho-

mologies. As discussed earlier, the virus-specified sequence for the three joints is perfectly homologous (Fig. 5 and 6b). Comparison of the flanking regions out to 80 nucleotides from the joint fails to reveal obvious homologies either between the viral DNA and host sequences or among the three host sequences. Partial homologies between the left end of the LTR and the 3′ host sequences can be found, particularly in the case of clones 2a1-4 and 3a1-3, but the extent of these structures is not compelling. Sequences from a large number of integration events will undoubtedly be needed to resolve the potential role of sequence homology in retroviral integration.

One other feature is to be noted in the adjacent host sequences (Fig. 6b). Clone 3a1-3 contains a 22-base direct repeat from position +35 to +56 and +57 to +78. Included within this repeat are the last seven bases (TGCGGCA) encoded by the integrated viral genome. Clone 4a1-1 also contains a 10-base direct repeat (position +54 to +63 and +69 to +78). Determination of whether these curious sequences actually play a role in the integration process again must await the accumulation of a larger number of integration site sequences.

From the information available thus far, we believe that it is unlikely that viral integration involves the recognition of a unique DNA sequence. Since we have only examined 3′ joints, the formal possibility remains that a potential recognition site could be asymmetrically displaced to the 5′ end of the integrated provirus. Evidence in other systems (44) in which both flanking regions have been compared indicates that this is not the case.

ACKNOWLEDGMENTS

We thank D. Berard, R. Wolford, and B. Fleurdelys for excellent technical assistance, L. Kunz for growth and purification of plasmids, and L. Johnson, M. Ostrowski, and D. Wheeler for helpful discussions. We are also grateful to Lorraine Shaughnessy for typing the manuscript. J. Majors and H. Diggelmann generously communicated data before publication.

This work was also supported in part by a contract from the National Cancer Institute to Meloy Laboratories, Springfield, Va.

LITERATURE CITED

1. **Baker, C. C., J. Herisse, G. Courtois, F. Galibert, and E. Ziff.** 1977. Messenger RNA for the Ad2 DNA binding protein: DNA sequences encoding the first leader and heterogeneity at the mRNA 5′ end. Cell **18**:569–580.
2. **Benton, D., and R. Davis.** 1977. Screening λgt recombinant clones by hybridization to single plaques in situ. Science **196**:180–182.
3. **Blattner, F. R., B. G. Williams, A. Blechl, K. Denniston-Thompson, H. E. Faber, L. Furling, D. J. Grunwald, D. O. Keifer, D. D. Moore, J. W. Schumm, E. L. Sheldon, and O. Smithies.** 1977. Charon phages: safer derivatives of bacteriophage lambda for DNA cloning. Science **196**:159–161.
4. **Brown, D. D., L. J. Korn, E. Birkenmeier, R. Peterson, and S. Sakonju.** 1979. The in vitro transcription of Xenopus 5S DNA, p. 511–519. *In* R. Axel, T. Maniatis, and C. F. Fox (ed.), Eucaryotic gene regulation, ICN-UCLA Symposia on Molecular and Cellular Biology, vol. 14.
5. **Calos, M. P., L. Johnsrud, and J. H. Miller.** 1978. DNA sequence at the integration sites of the insertion element IS1. Cell **13**:411–418.
6. **Cohen, J. C., P. R. Shank, V. L. Morris, R. Cardiff, and H. E. Varmus.** 1979. Integration of the DNA of mouse mammary tumor virus in virus-infected normal and neoplastic tissue of the mouse. Cell **16**:333–345.
7. **Cohen, J. C., and H. E. Varmus.** 1979. Endogenous mammary tumor virus DNA varies among wild mice and segregates during inbreeding. Nature (London) **278**:418–423.

7a. **Dhar, R., W. L. McClements, L.W. Enquist, and G. F. Vande Woude.** 1980. Nucleotide sequences of integrated Moloney sarcoma provirus long terminal repeats and their host and viral junctions. Proc. Natl. Acad. Sci. U.S.A. **77**:3937–3941.

7b. **Donehower, L. A., J. Andre, D. S. Berard, R. G. Wolford, and G. L. Hager.** 1980. Construction and characterization of molecular clones containing integrated mouse mammary tumor virus sequences. Cold Spring Harbor Symp. Quant. Biol. **44**:1153–1159.

8. **Fine, D. L., L. O. Arthur, J. K. Plowman, E. A. Hillman, and F. Klein.** 1974. In vitro system for production of mouse mammary tumor virus. Appl. Microbiol. **28**:1040–1046.
9. **Gannon, F., K. O'Hare, F. Perrin, J. P. LePennec, C. Benoist, M. Cochett, R. Breathnach, A. Royal, A. Garapin, B. Cami, and P. Chambon.** 1979. Organisation and sequences at the 5′ end of a cloned complete ovalbumin gene. Nature (London) **278**:428–434.
10. **Grindley, N. D. F.** 1978. IS1 insertion generates duplication of a nine base pair sequence at its target site. Cell **13**:419–426.
11. **Gross-Bellard, M., P. Oudet, and P. Chambon.** 1973. Isolation of high molecular weight DNA from mammalian cells. Eur. J. Biochem. **36**:32–38.
12. **Grunstein, M., and D. Hogness.** 1975. Colony hybridization; a method for the isolation of cloned DNAs that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. **72**:3961–3965.
13. **Hager, G. L., E. H. Chang, H. W. Chan, C. F. Garon, M. A. Israel, M. A. Martin, E. M. Scolnick, and D. R. Lowy.** 1979. Molecular cloning of the Harvey sarcoma virus closed circular DNA intermediates: initial structural and biological characterization. J. Virol. **31**: 795–809.
14. **Hagerbuchle, O., M. Santer, and J. Argetsinger Steitz.** 1978. Conservation of the primary structure at the 3′ end of 18S rRNA from eucaryotic cells. Cell **13**: 551–563.
15. **Haseltine, W. A., and D. G. Kleid.** 1978. A method for classification of 5′ termini of retroviruses. Nature (London) **273**:358–364.
16. **Hilgers, J., and P. Bentvelzen.** 1978. Interaction between viral and genetic factors in murine mammary cancer. Adv. Cancer Res. **26**:143–195.
17. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
18. **Hsu, T. W., J. L. Sabron, G. E. Mark, R. V. Guntaka, and J. M. Taylor.** 1978. Analysis of unintegrated RNA tumor virus double-stranded DNA intermediates. J. Virol. **28**:810–818.
19. **Hughes, S. H., P. R. Shank, D. H. Spector, H.-J. Kung, J. M. Bishop, H. E. Varmus, P. K. Vogt, and M. L. Breitman.** 1978. Proviruses of avian sarcoma

virus are terminally redundant, co-extensive with unintegrated linear DNA and integrated at many sites. Cell **15**:1397–1410.

20. **Jeffreys, A. J., and R. A. Flavell.** 1977. A physical map of the DNA regions flanking the rabbit β-globin gene. Cell **12**:429–439.
21. **Konkel, D. A., J. V. Maizel, and P. Leder.** 1979. The evolution and sequence comparison of two recently diverged mouse chromosomal β-globin genes. Cell **18**: 865–873.
22. **Korn, L. J., and D. D. Brown.** 1978. Nucleotide sequence of *Xenopus borealis* oocyte 5S DNA: comparison of sequences that flank several related eucaryotic genes. Cell **15**:1145–1156.
23. **Lai, E. C., J. P. Stein, J. F. Cotterall, S. L. C. Woo, M. L. Mace, A. R. Means, and B. W. O'Malley.** 1979. Molecular structure and flanking nucleotide sequences of the natural chicken ovomucoid gene. Cell **18**:829–842.
24. **Landy, A., and W. Ross.** 1977. Viral integration and excision: structure of the lambda att sites. Science **197**: 1147–1160.
25. **Leder, P., D. Tiemeier, and L. Enquist.** 1977. EK2 derivatives of bacteriophage lambda useful in the cloning of DNA from higher organisms: the λgtWES system. Science **196**:175–177.
26. **Lerner, M. R., J. A. Boyle, S. M. Mount, S. L. Wolin, and J. A. Steitz.** 1980. Are snRNPs involved in splicing? Nature (London) **283**:220–224.
27. **Mandel, M., and A. Higa.** 1970. Calcium-dependent bacteriophage DNA infection. J. Mol. Biol. **53**:159–162.
28. **Maniatis, T. R., R. C. Hardison, E. Lacy, U. Laver, C. O'Connell, D. Quon, G. K. Sim, and A. Efstradiadis.** 1978. The isolation of structural genes from libraries of eucaryotic DNA. Cell **15**:687–701.
29. **Manley, J. L., P. A. Sharp, and M. L. Gefter.** 1979. RNA synthesis in isolated nuclei: *in vitro* initiation of adenovirus 2 major late mRNA precursor. Proc. Natl. Acad. Sci. U.S.A. **76**:160–164.
30. **Maxam, A., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. U.S.A. **74**:560–564.
31. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
32. **Nishioka, Y., and P. Leder.** 1979. The complete sequence of a chromosomal mouse α-globin gene reveals elements conserved throughout vertebrate evolution. Cell **18**:875–882.
33. **Owens, R. B., and A. J. Hackett.** 1972. Tissue culture studies of mouse mammary tumor cells and associated viruses. J. Natl. Cancer Inst. **49**:1321–1328.
34. **Parks, W. P., E. S. Hubbell, R. J. Goldberg, F. J. O'Neill, and E. M. Scolnick.** 1976. High frequency variation in mammary tumor virus expression in cell culture. Cell **8**:87–93.
35. **Pribnow, D.** 1975. Bacteriophage T7 early promoters: nucleotide sequences of two RNA polymerase binding sites. J. Mol. Biol. **99**:419–443.
36. **Ringold, G., K. R. Yamamoto, P. R. Shank, and H. E. Varmus.** 1977. Mouse mammary tumor virus DNA in infected rat cells: characterization of unintegrated forms. Cell **10**:19–26.
37. **Ringold, G. M., P. R. Shank, H. E. Varmus, J. Ring, and K. R. Yamamoto.** 1979. Integration and transcription of mouse mammary tumor virus DNA in rat hepatoma cells. Proc. Natl. Acad. Sci. U.S.A. **76**:665–669.
38. **Rosenberg, M., and D. Court.** 1979. Regulatory sequences involved in the promotion and termination of RNA transcription. Annu. Rev. Genet. **13**:319–353.
39. **Sabran, J. L., T. W. Hsu, C. Yeater, A. Kaji, W. S. Mason, and J. M. Taylor.** 1979. Analysis of integrated avian RNA tumor virus DNA in transformed chicken, duck, and quail fibroblasts. J. Virol. **29**:170–178.
40. **Sanger, F., and A. R. Coulson.** 1978. The use of thin acrylamide gels for DNA sequencing. FEBS Lett. **87**: 107–110.
41. **Shank, P. R., J. C. Cohen, H. E. Varmus, K. R. Yamamoto, and G. M. Ringold.** 1978. Mapping of linear and circular forms of mouse mammary tumor virus DNA with restriction endonucleases: evidence for a large specific deletion occurring at high frequency during circularization. Proc. Natl. Acad. Sci. U.S.A. **75**: 2112–2116.
42. **Shank, P. R., S. H. Hughes, H. S. Kung, J. E. Majors, N. Quintrell, R. V. Guntaka, J. M. Bishop, and H. E. Varmus.** 1978. Mapping unintegrated avian sarcoma virus DNA: termini of linear DNA bear 300 nucleotides present once or twice in two species of circular DNA. Cell **15**:1383–1395.
43. **Shimotohne, K., S. Mijutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**: 550–554.
44. **Shoemaker, C., S. Goff, E. Gilboa, M. Paskind, S. W. Mitra, and D. Baltimore.** 1980. Structure of a cloned circular Moloney murine leukemia virus DNA molecule containing an inverted segment: implications for retrovirus integration. Proc. Natl. Acad. Sci. U.S.A. **77**:3932–3936.
45. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **38**:503–517.
46. **Stauffer, G. V., G. Zurawski, and C. Yanofsky.** 1978. Single base-pair alterations in the *Escherichia coli trp* operon leader region that relieve transcription termination at the *trp* attenuator. Proc. Natl. Acad. Sci. U.S.A. **75**:4833–4837.
47. **Taylor, J. M., R. Ilmensee, and J. Summers.** 1976. Efficient transcription of RNA into DNA by avian sarcoma virus polymerase. Biochim. Biophys. Acta **442**: 324–330.
48. **Thompson, E. B., D. Aviv, and M. E. Lippman.** 1977. Variants of HTC cells with low tyrosine aminotransferinase inducibility and apparently normal glucocorticoid receptors. Endocrinology **100**:406–419.
49. **Thompson, E. B., D. K. Granner, T. Gelehrter, J. Erickson, and G. L. Hager.** 1979. Unlinked control of multiple glucocorticoid-induced processes in HTC cells. Mol. Cell. Endocrinol. **15**:135–150.
50. **Tsujimoto, Y., and Y. Suzuki.** 1979. Structural analysis of the fibrion gene at the 5′ end and its surrounding regions. Cell **16**:425–436.
51. **Tu, C., and S. N. Cohen.** 1980. Translocation specificity of the Tn3 element: characterization of sites of multiple insertions. Cell **19**:151–160.
52. **Weil, P. A., D. S. Luse, J. Segall, and R. G. Roeder.** 1979. Selective and accurate initiation of transcription at the Ad2 major late promoter in a soluble system dependent on purified RNA polymerase II and DNA. Cell **18**:469–484.
53. **Yamamoto, T., G. Jay, and I. Pastan.** 1980. Unusual features in the nucleotide sequence of a cDNA clone derived from the common region of avian sarcoma virus messenger RNA. Proc. Natl. Acad. Sci. U.S.A. **77**:176–180.
54. **Ziff, E. B., and R. M. Evans.** 1978. Coincidence of the promoter and capped 5′ terminus of RNA from the adenovirus 2 major late transcription unit. Cell **15**: 1463–1475.