MOLECULAR AND CELLULAR BIOLOGY, Dec. 1983, p. 2221–2231
0270-7306/83/122221-11$02.00/0

Vol. 3, No. 12

Structure and Expression of Mouse VL30 Genes

CLAGUE P. HODGSON,† PAULA K. ELDER, TETSUYA ONO,‡ DOUGLAS N. FOSTER, AND MICHAEL J. GETZ*

Department of Cell Biology, Mayo Clinic/Foundation, Rochester, Minnesota 55905

Received 15 June 1983/Accepted 13 September 1983

DNA sequencing and blot hybridization analyses have been used to study the structure of a mouse VL30 gene and the molecular nature of VL30-related RNA which is induced upon the stimulation of cultured AKR mouse embryo cells with defined peptide growth factors. An integrated mouse VL30 gene was found to contain identical 601-base-pair long terminal repeats (LTRs) which were themselves terminated in short inverted repeats. The entire VL30 gene was flanked by a 4-base-pair direct repeat of cellular DNA. Thus, VL30 genes are structurally analogous to integrated forms of retrovirus proviruses and certain other classes of mobile genetic elements. The LTR sequence was found to contain putative promoter and polyadenylation signals and generally exhibited little sequence homology to murine leukemia virus proviral LTRs. Certain short regions of sequence conservation, however, were evident, including the inverted terminal repeat, LTR-adjacent regions corresponding to origins of murine leukemia virus proviral DNA synthesis, and a 36-base-pair direct repeat bearing homology to the 72-base-pair direct repeat (enhancer sequence) of the murine leukemia virus-related Moloney sarcoma virus. Upon mitogenic stimulation of quiescent cells with epidermal growth factor and insulin, a major 5.5-kilobase VL30-specific RNA complementary to both LTR and non-LTR sequences was rapidly induced. We conclude that a complete VL30 gene(s) is highly regulated by peptide growth factor binding to specific membrane receptors in these cells.

Mouse VL30 elements are a family of dispersed 5- to 6-kilobase (kb) DNA sequences present in 100 to 200 similar copies per genome (19, 21). The predominant transcription product is a 30S polyadenylated RNA which is expressed in a variety of cell types (2, 6, 17). VL30 RNA encodes no known polypeptides but is copackaged with high efficiency into the virions of replication-competent retroviruses (35) and can be transmitted to heterologous cell types in culture (33). It is not known whether VL30 elements can replicate independently of helper retroviruses.

VL30 sequences appear to have arisen in rodent genomes before *Mus* speciation, as evidenced by their presence in all tested representatives of the genus *Mus* as well as in rats (1, 5). However, considerable locus polymorphism can be distinguished in different laboratory strains and feral individuals. We have previously determined that one such polymorphism reflects the differing chromosomal location of a complete or nearly complete VL30 element (5). Therefore, VL30 sequences appear to be mobile, at least over evolutionary time spans. Consistent with this property, heteroduplex and blot hybridization analyses have demonstrated that at least some VL30 elements contain terminal redundancies similar to the long terminal repeats (LTRs) of retrovirus proviruses and certain other classes of transposable genetic elements (20). Thus, although the precise nature and physiological significance of this multigene family are not known, current evidence suggests that VL30 genes represent either defective retrovirus genomes acquired early in rodent evolution or a class of cellular mobile elements which have acquired the ability to be efficiently packaged into retrovirus virions.

Recent studies from our laboratory have shown that the abundance of VL30-related RNA in cultured AKR mouse embryo cells is highly dependent on the binding of peptide growth factors to membrane receptors (11). In contrast, endogenous murine leukemia virus (MuLV) proviruses are not differentially regulated by peptide growth factors, suggesting that VL30 genes possess regulatory properties unique to this class of elements.

In the present study, we have examined the nucleotide sequence of the terminally redundant

† Present address: Department of Cell Biology, Baylor College of Medicine, Houston, TX 77970.
‡ Present address: Department of Radiation Biophysics, Faculty of Medicine, University of Tokyo, Tokyo, 113, Japan.

regions and associated flanking sequences of an integrated mouse VL30 gene. These studies establish the presence of structural features generally regarded as hallmarks of mobile genetic elements. In addition, we have examined the molecular characteristics of VL30-related RNA in peptide growth factor-stimulated cells. This RNA appears to be a complete transcript of an integrated VL30 gene(s).

(A portion of this work constituted a Ph.D. dissertation presented to the faculty of the University of Minnesota by C.P.H.)

MATERIALS AND METHODS

Cell culture. Mouse embryo-derived AKR-2B cells were cultured as previously described (11) with the following modifications. Nearly confluent cultures were rinsed once with serum-free, unsupplemented medium MCDB 402 (K. C. Biological, Lenexa, Kans.) and incubated in the same medium for 48 h (G. Shipley, submitted for publication). Cells were stimulated to proliferate by the addition of epidermal growth factor (EGF) (10 ng/ml) and insulin (500 ng/ml) in 10 mM glucose–3 mM KCl–130 mM NaCl–1.0 mM $Na_2HPO_4 \cdot 7H_2O$–0.0033 mM phenol red–30 mM HEPES (*N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid), pH 7.6. Cells were harvested at 0, 2, 4, and 6 h after stimulation. The growth rates of all experimental cultures were determined both before and 20 h after stimulation by measuring the rate of [^{3}H]thymidine incorporation as previously described (13). Cell fractionation and purification of polysomal RNA have been described previously (13).

Construction of subclones. Subclones of the BVL-1 locus and flanking cellular regions were constructed with plasmids pBR322 and pACYC177, which had been restriction enzyme digested either with *Eco*RI and *Hin*dIII or with *Xho*I, *Sma*I, or *Pst*I, respectively, depending on the particular fragment to be subcloned. Corresponding BVL-1 fragments were digested, electrophoresed through 0.8% agarose gels, and eluted. Ligation was performed at a 1:1 molar ratio of vector to target DNA at a concentration of 0.5 to 1 μg/μl in a 1× ligase buffer (20 mM Trizma base [pH 7.3], 7.5 mM $MgCl_2$, 0.1 mM EDTA, 0.5 mM ATP) with a ≤0.1 volume of ligase in plastic tubes at 15°C. After 15 min, a sample of the ligation mixture was electrophoresed to monitor the extent and forms of ligation produced. Ligations were stored at 4°C until used.

Northern blot analysis of RNA. Electrophoresis of RNA in 1% agarose gels containing formaldehyde and formamide was performed exactly as described by Maniatis et al. (28). After electrophoresis, gels were soaked sequentially for 1 h each at room temperature in (i) several changes of distilled water; (ii) 50 mM NaOH–10 mM NaCl; (iii) 100 mM Tris-hydrochloride, pH 7.5; and (iv) 20× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate, pH 7.0). The RNA was transferred to presoaked nitrocellulose filter paper by blotting overnight at room temperature in 20× SSC. Filters were then air dried and baked for 2 to 4 h at 80°C in a vacuum oven. Before hybridization, filters were washed for 15 to 30 min in 3× SSC at room temperature followed by incubation in hybridization buffer for 2 to 4 h at 43°C in sealed plastic bags (hybridization buffer is 5× SSC containing 50% formamide, 250 μg of sonicated herring sperm DNA per ml, 50 μg of polyadenylic acid per ml, 1× Denhardt solution [0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% bovine serum albumin], 0.1% sodium dodecyl sulfate). The appropriate ^{32}P-labeled probe (0.5 μg), nick translated by using a commercially available kit (New England Nuclear Corp., Boston, Mass.), was added, and the incubation was continued for 16 to 20 h at 43°C. Filters were then given three quick rinses in 5× SSC at room temperature, followed by washes in 5× SSC for 30 min at 43°C, and in fresh hybridization buffer for 1 to 2 h at 43°C. Filters were air dried and autoradiographed by exposure to Kodak X-Omat X-ray film.

DNA sequencing. Sequencing was carried out using the unmodified protocols of Maxam and Gilbert (29). End-labeling reactions were performed by polynucleotide kinase treatment of 5'-dephosphorylated ends with [γ-^{32}P]ATP or by adding cordycepin (2,3-dideoxyadenosine triphosphate, α-^{32}P labeled) to 3' ends with terminal transferase (36). Both sets of reactions were carried out using the kits and protocols supplied by the manufacturer (New England Nuclear Corp.). Sequencing gels (0.8 mm) were run at 2,000 to 2,500 V at room temperature with 8 or 20% acrylamide, 8 M urea, and a 20:1 ratio of acrylamide to bisacrylamide. Gels were either 40 cm (for 20% gels) or 88 cm. Sequencing gels were transferred to X-ray film and generally were exposed overnight at −80°C.

Computer analysis of DNA sequence data. The various programs used in sequence analysis were provided through a paid time-sharing agreement with Intelligenetics, Inc., Palo Alto, Calif. and have been described in detail elsewhere (4). Occasionally, sequences were searched from the data banks of the EMBO (European Molecular Biology Organization) or Los Alamos storage facilities, and in all cases these were accessed directly via Intelligenetics.

RESULTS

Terminally redundant sequences in different VL30 clones. Figure 1 illustrates restriction endonuclease maps for three different VL30 clones. Clone BVL-1 was isolated from a BALB/c genomic library (5), whereas clones AKVL-8 and AKVL-27 were isolated from an AKR-2B cell genomic library. The hatched areas represent presumptive LTRs based on the fact that sequences in these regions all hybridized to a subcloned probe corresponding to the left terminus of clone BVL-1 (data not shown). One of the clones, AKVL-27, hybridized to the terminal region probe only at one end, reflecting the fact that this clone contains an incomplete VL30 gene.

Nucleotide sequence of VL30 LTRs. The sequences of both terminally redundant ends of clone BVL-1 were determined by the techniques of Maxam and Gilbert (29) on subcloned segments of the locus (Fig. 2). Generally, the sequence was determined for the 5'-end-labeled strand more than once, or for the 5'-labeled strands in both directions, or for both the 5'- and

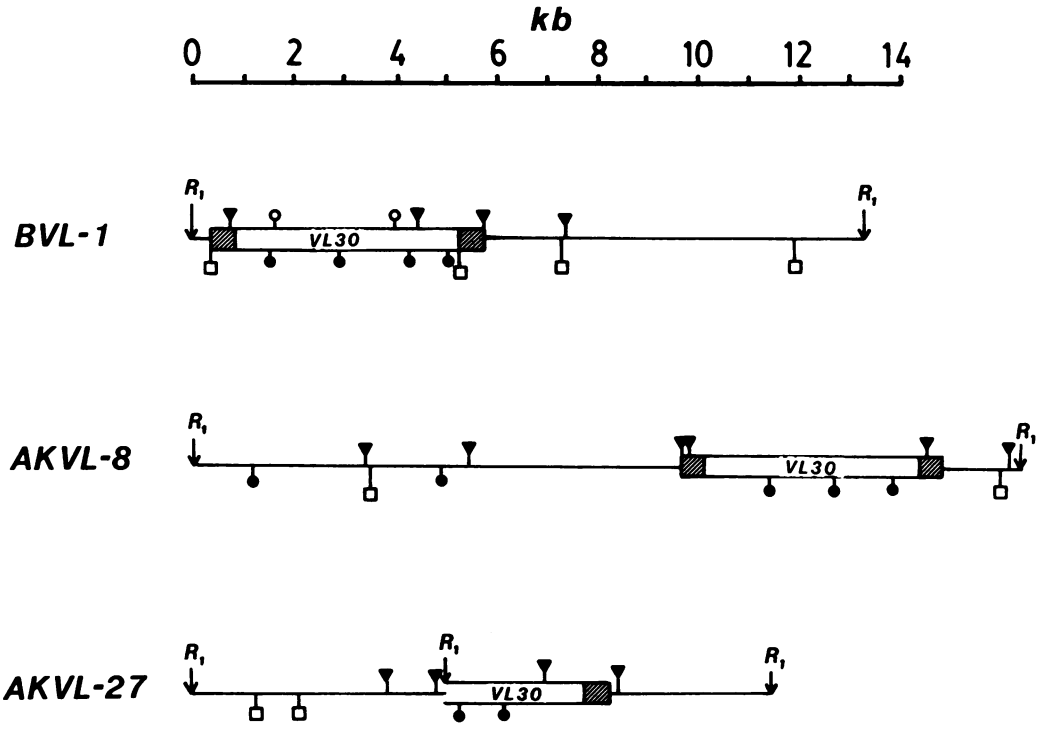

FIG. 1. Restriction enzyme maps of three Charon 4A clones of VL30 elements and flanking cellular sequences. Hatched regions represent presumptive LTRs based on hybridization to a terminal region probe derived from BVL-1 (see text). Restriction enzyme cleavage sites are *Eco*RI (R_1), *Xba*I (▼), *Xho*I (□), *Hin*dIII (●), and *Pst*I (○).

3'-end-labeled strands. All restriction site junctions used in labeling were sequenced across. Altogether, over 95% of the sequences were determined at least twice on each LTR. In the regions where an LTR was sequenced once, no differences were encountered between the LTRs.

The boundaries of the terminal repeats were assigned by the points of divergence between the two sequences. By this criterion, the length of the LTR sequence was determined to be 601 base pairs (bp), and no differences were encountered between the two LTRs. It should be noted, however, that two nucleotides are characteristically lost from each end of linear retrovirus proviral DNA upon integration. The VL30 LTR sequence determined here can also be interpreted as having resulted from the loss of the dinucleotide AA from one junction with flanking cellular DNA and the loss of the complementary dinucleotide TT from the opposite junction. If this is the case, the actual length of the integrated LTR sequence would be 603 bp, and the two LTRs would differ by a dinucleotide pair (TT versus AA) at the internal junctions with the remainder of the VL30 unit. Establishing this point, however, would require sequencing of the unintegrated VL30 DNA precursor. The complete sequence, including flanking element and cellular regions, is shown in Fig. 3. The major structural features (see below) are indicated in the Figure. Bordering the point of sequence divergence inside both termini of the two LTRs was an inverted repeat of 4 bp, TGAA . . . TTCA.

A 4-bp direct repeat of DNA (GCTG . . . GCTG) was found congruent with the predicted ends of the VL30 sequence. Although we have not yet sequenced through an unoccupied acceptor site similar to a retroviral preinsertion site in cellular DNA, we assume that this short direct repeat resulted from the duplication of a cellular sequence during insertion of the VL30 element into the host chromosome.

Sequence homology to LTRs of known murine retroviruses. Although it was previously believed that MuLVs and mouse VL30s are not genetically related (2, 17, 35), recent results with reduced-stringency hybridization conditions demonstrate distinct, overlapping regions of homology between rat and mouse VL30s and MuLVs (14). None of these regions detectable by blot hybridization include the LTRs. Direct comparison of sequence data, however, revealed distinct homology in a short region between the Moloney MuLV (MoMuLV) minus-

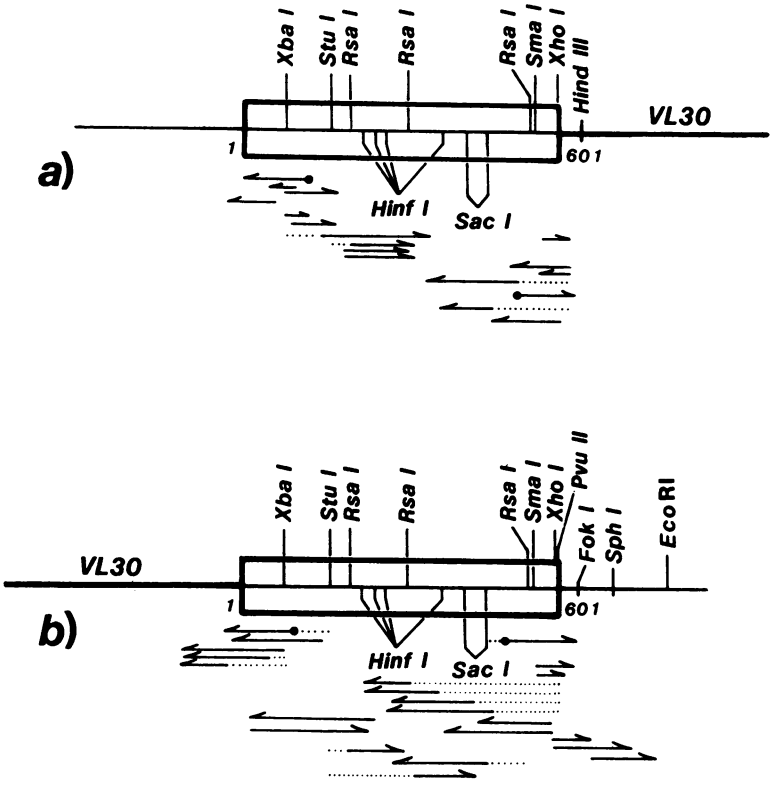

FIG. 2. Strategy for sequencing the LTR regions of mouse VL30 clone BVL-1. Restriction enzyme maps of the two LTRs, showing sequencing strategy for each region. Straight arrows represent 5'-end-labeled fragments, and arrows with circles represent 3'-end-labeled fragments. (a) Right LTR; and (b) left LTR. Right and left orientation indicate position in the Charon 4A clone as shown in Fig. 1. Numbering goes from right to left.

strand primer binding site (PBS) and an analogous region bordering one edge of the VL30 LTR (Fig. 4A). A 15-bp perfect homology overlaps the end of the LTR and extends into the PBS, with imperfect homology thereafter. As previously noted (14), the homology extending into the first 10 bp of the LTR corresponds exactly to the first 10 nucleotides of an 11-bp inverted repeat sequence found at the termini of all integrated forms of MuLV proviruses thus far examined (38). The fact that this short but significant homology overlaps the ends of VL30 and MuLV LTRs, combined with the fact that the predicted ends of the VL30 agree to the exact nucleotide with MuLV ends, further strengthens the idea that the ends of the VL30 LTR predicted by the points of divergence are accurate with respect to the integrated sequence. Finally, an 18-bp complementarity (up to the first modified base) to a $tRNA^{Gly}$ 3' end was found at the same position in which $tRNA^{Pro}$ binds in the analogous retroviral minus-strand PBS sequence (Fig. 4A). Interestingly, a region showing a similar homology to the plus-strand origin of MuLV replication was found bordering the opposite LTR (Fig. 4B). Beyond the inverted repeats and adjoining VL30 regions, little homology was found between the VL30 LTR and the LTRs of either MoMuLV, mouse mammary tumor virus (MMTV), or the mobile dispersed genetic element 1 or *copia* transposable elements of *Drosophila*. Two regions of imperfect homology to MuLVs, however, are shown in Fig. 5. One region (Fig. 5B) overlaps a set of 36-bp direct repeats in the VL30 LTR and has homology to a region found in a set of 72-bp direct repeats in the U3 region of the MuLV-related Moloney murine sarcoma virus (MoMSV) known as an enhancer sequence because it has the ability to enhance transcription of RNA from nearby promoters (27).

Molecular nature of peptide growth factor-induced VL30 RNA. Although previous studies have shown that VL30-related RNA is rapidly induced after mitogenic stimulation of quiescent AKR-2B cells with EGF and insulin, the molecular nature of this RNA has not been determined. Thus, it could represent either a complete VL30 gene transcript, the transcription product of a related cellular gene, or a hybrid

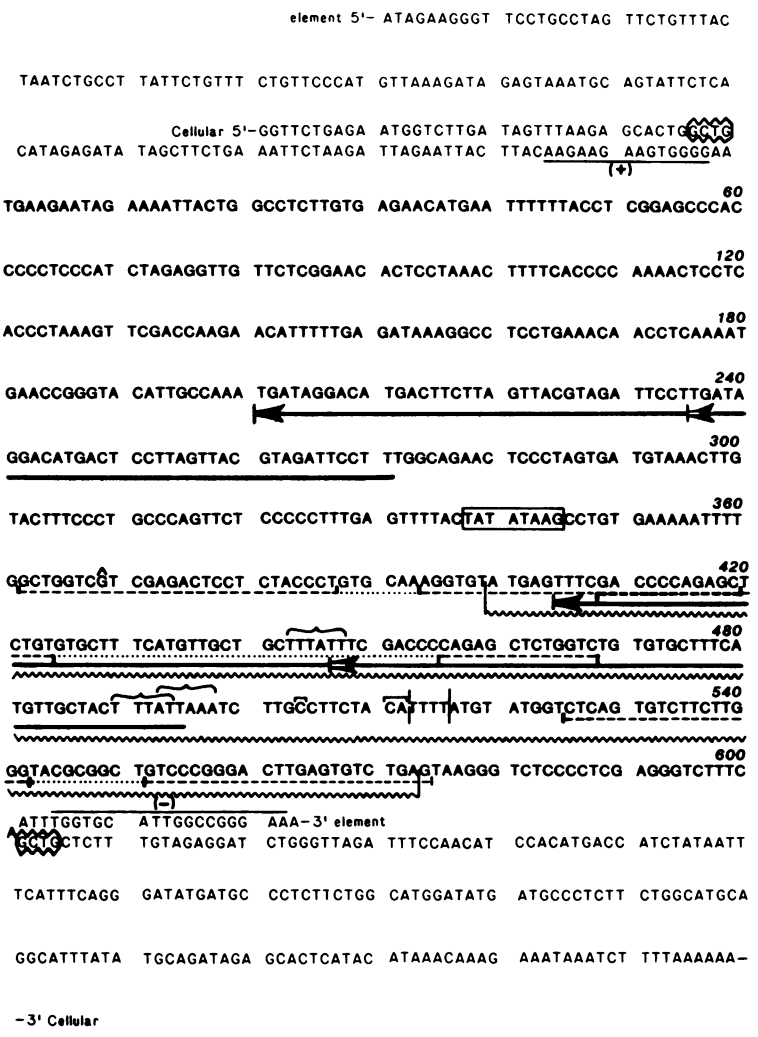

FIG. 3. Nucleotide sequence of VL30 LTRs and flanking sequences. LTR sequences are bold, and numbering starts at the first nucleotide of the LTR; flanking sequences are in light lettering. The terminal direct repeats of cell DNA are enclosed in scallops, internal direct repeats are underlined in bold arrows, regions of dyad symmetry are indicated by dashed underlining with dots where loop regions of potential hairpins are located, and overlapping regions of dyad symmetry are shown by bold dashed underlining. Potential transcriptional signals include the following: TATATAAG box (enclosed), polyadenylation-like signal (raised brace), polyadenylation site (⌐), T cluster (possible terminal signal) (parallel vertical lines), open reading frame (wavy underline), PBS (light underline), possible TATA boxes in opposite orientation (lowered braces), preferred retroviral cap site (G).

RNA resulting from downstream promotion from a VL30 LTR. To investigate this issue, several subcloned probes were prepared from λ clone BVL-1. A 401-bp *Stu*I-*Sma*I fragment, entirely internal to the LTR sequence (see Fig. 2), was subcloned into the *Sma*I site of pACYC177. The resulting plasmid, termed pVLTR4, was used as an LTR-specific probe. A second plasmid, termed pVL17, was constructed by subcloning a 1.7-kb internally derived *Pst*I fragment (Fig. 1) into pBR322. This plasmid contained no LTR sequences. Finally, an approximately 5-kb *Xho*I fragment (Fig. 1) was subcloned into the *Xho*I site of pACYC177. This plasmid, termed pVL47, contained a complete VL30 gene with the exception of one copy of the LTR.

Quiescent mouse embryo-derived AKR-2B cells in serum-free medium were stimulated to proliferate by the addition of EGF and insulin. Polysomal RNA was isolated at various times after stimulation and analyzed for VL30 RNA by Northern blot hybridization by using the subcloned probes. Figure 6A illustrates that the

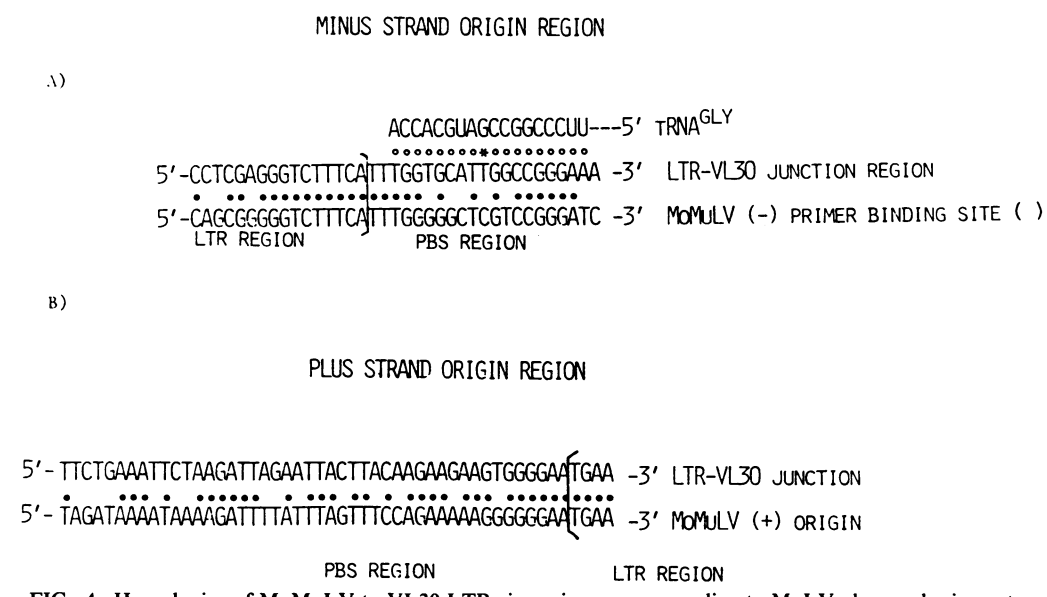

FIG. 4. Homologies of MoMuLV to VL30 LTRs in regions corresponding to MuLV plus- and minus-strand origins of DNA replication. (A) Minus-strand origin region of MoMuLV (37) showing the end of the integrated LTR and adjacent PBS region, as well as the partially homologous region of BVL-1. A $tRNA^{Gly}$ (12) 3' terminus which is complementary for 18 bp to the VL30 in this region is also shown up to the first modified base. Symbols: (●), homology; (○), complementarity; and *, potential GU pair. (B) Plus-strand origin region of MoMuLV showing the LTR terminus and its VL30 counterpart.

complete probe (pVL47) detected a major inducible RNA species with an apparent chain length of 5.5 kb. Figure 6B illustrates that a similar 5.5-kb RNA can be detected with both the LTR (pVLTR4) and the internal sequence (pVL17) probe. Thus, the major EGF-inducible RNA in these cells is similar in size to an integrated VL30 gene and appears to contain both LTR and non-LTR-specific information.

Although a number of minor bands were observed when internally derived (non-LTR) sequences were present in the hybridization probe, their significance is open to question. Since they were not detected with the LTR probe, it is possible that they represent transcription products of related genes which lack LTR information. We have been unable to detect such sequences in genomic DNA, however, and consider it more likely that they correspond to degradation products. The apparent lack of LTR information in these minor species may be misleading, since recent data from our laboratory and others (E. Keshet, personal communication) suggest considerable LTR sequence divergence among the family of VL30 units. Although we cannot presently estimate what fraction of these units are transcribed upon EGF stimulation, these data suggest that the major transcription products correspond to full-length copies of at least one complete VL30 gene.

DISCUSSION

These data extend the results of Keshet and Shaul (20) which demonstrate that at least some VL30 elements contain terminally redundant sequences. Moreover, these data provide direct confirmation that such terminally redundant sequences display several structural features which are characteristic of integrated forms of retrovirus proviruses and certain other classes of transposable genetic elements. The VL30 element in clone BVL-1 has been shown to contain identical 601-bp LTRs which are them-

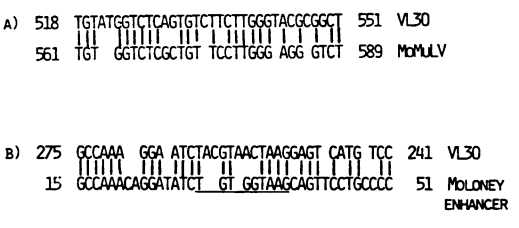

FIG. 5. Regions of limited homology between MoMuLV and VL30 LTRs. (A) Homology between a U5 region in MoMuLV (38) and a similar sequence in BVL-1 LTRs. (B) Homology between Moloney enhancer sequences found on a 72-bp direct repeat of MoMSV (27) and a sequence found overlapping a set of 36-bp repeats in the VL30 LTRs.

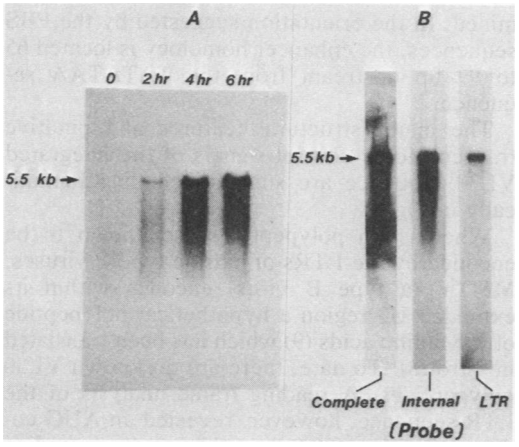

FIG. 6. VL30-specific RNA species induced by stimulation of quiescent AKR-2B cells with EGF and insulin. (A) Each lane contained 4 μg of total polysomal RNA from cells harvested at the indicated times after stimulation. (B) Total polysomal RNA from cells 6 h after stimulation probed with the indicated subcloned probes as described in the text. The chain length was estimated with reference to 16S, 18S, 23S, and 28S rRNA markers coelectrophoresed in a parallel lane.

selves terminated in short inverted repeats. The sequence of these inverted repeats, TGAA . . . TTCA, results in the same beginning and ending dinucleotide pair displayed by many transposable elements and all integrated proviruses thus far examined (39). In addition, the entire VL30 element was found to be flanked by a 4-bp direct repeat of cell DNA. It seems likely, though not yet proven, that this short direct repeat results from the duplication of a single copy of this sequence during the insertion of the VL30 element into the host chromosome. These circumstances, LTRs bounded by a short, direct repeat of cell DNA and an inverted repeat sequence ending in TG . . . CA at the ends of element DNA, are also the hallmarks of many other cellular and retroviral mobile genetic elements.

These data also suggest a molecular basis for the notion that VL30 elements replicate and disperse themselves via MuLV-assisted mechanisms. First, the direct repeat of DNA found flanking the terminal repetition (4 bp) was the same size as the direct repeat created by the integration of most MuLV proviruses (7), suggesting that mechanisms similar to those of the MuLVs may be involved in the insertion into the genome of the VL30 element. Second, the terminal repeats at the ends of the 601-bp sequences consist of the same 4 bp as are found at the ends of MuLVs (38). In addition, the sequence at one terminus of the LTR is homologous to MuLV for the first 10 bp of the 11-bp MuLV terminal inverted repeat. Third, this same terminal 10-bp homology to MuLV is adjacent to a non-LTR VL30 sequence which bears homology to the MoMuLV minus-strand PBS. Fourth, the VL30 has an 18-bp complementarity to the 3′ end of a $tRNA^{Gly}$ starting at the site analogous to the minus-strand PBS. Fifth, the opposite VL30-LTR junction has homology to the MoMuLV plus-strand origin of replication.

One interpretation of these findings is that both the integration and replication of VL30 elements proceed by the same or similar mechanisms as do MuLVs. A strong selection pressure to utilize MuLV-encoded mechanisms could account for the relative conservation of the MuLV origins of DNA synthesis and the inverted terminal repeat. Such inverted terminal repeats are believed to function in promoting the integration of mobile genetic elements into the host chromosome, possibly via a site-specific endonuclease activity (34).

The sequence data presented here suggest that at least some VL30 LTRs may provide control signals for regulating the synthesis of VL30 RNA. Furthermore, the observed homologies to MuLV plus- and minus-strand origins of DNA synthesis permit, by analogy to the direction of MuLV transcription, speculation as to the orientation of transcription of the cloned VL30 element studied here (for a review of retrovirus transcription see Varmus [39]).

In the VL30 gene, the orientation of the PBS homologies suggests that transcription begins in the right LTR (upstream from the minus-PBS), proceeds leftward, and terminates in the left LTR. Directionality in this case refers only to the orientation of the cloned insert with respect to the arms of the original lambda vector and has no other significance. For ease of reference, the numbering and orientation of the LTR sequence is, therefore, right to left with respect to BVL-1. Several potential control signals are found in this orientation. These are as follows.

(i) A sequence resembling a TATA or Goldberg-Hogness box located at positions 338 through 344 (TATATAA). A similar sequence is frequently found about 23 to 31 bp upstream from the transcriptional start site of many eucaryotic genes (22, 24, 31). TATA boxes have been associated with transcriptional efficiency in vitro and may be responsible for the selection of the correct point of transcription initiation in vivo (3, 40).

(ii) The cap site, usually 24 nucleotides (± 1 bp) downstream from the TATA box 3′ end, is tentatively assigned to a position near the G (G is a preferred retroviral cap site) at position 369. This G is found at the center of a region of differential dyad symmetry (symmetry which would be energetically favored to assume differ-

ent secondary structure on RNAs containing either the 5′ or the 3′ ends) similar to the cap region of MoMuLV. This is the only such region of overlapping dyad symmetry found in the LTR, and it also overlaps the set of 52-bp direct repeats.

(iii) A polyadenylation signal-like sequence (ATTAAA) is located at positions 493 through 498 and is similar to the AATAAA sequence usually found 10 to 20 nucleotides upstream from many eucaryotic polyadenylation sites (10, 30, 32) and to the AGTAAA found in a similar position in the MMTV genome (9). Although the ATTAAA sequence is not common in eucaryotic mRNAs, it has been found near the polyadenylation site of α-amylase mRNA (15), and, most recently, rat γ-casein mRNA (16). Most significantly, the sequence TCTTGC which forms the 3′ end of mature MoMuLV viral RNA (7) was found immediately after the ATTAAA.

(iv) A T cluster (TTTT) which is often found about 8 nucleotides downstream from a polyadenylation site (22–24) is found 9 nucleotides from the TCTTGC sequence mentioned above as a possible polyadenylation site. All of the above sites can be seen in Fig. 3.

Other potential TATA-like or polyadenylation signal-like sequences are found in the direction opposing the orientation suggested above. The present data do not rule out their use in controlling element transcription.

The observed homology between a pair of 72-bp tandem repeats (enhancer sequence) found in the MoMSV and the 36-bp VL30 tandem direct repeat found at positions 241 through 275 suggests that VL30 LTRs may contain additional control regions which function in regulating the efficiency of transcription. The Moloney sequence appears functionally analogous to a set of 72-bp repeats found near the origin of replication of the simian virus 40 DNA virus (25). Although there is little actual sequence homology among the enhancer elements of simian virus 40, MoMSV, and several other viruses, Laimins et al. (25) have recently reported the existence of a set of core nucleotides which frequently occur in enhancer-type sequences of diverse origins. It is interesting to note that the MoMSV core sequence (TG$^{T}_{C}$GGTAAG) is located at the approximate center of the region exhibiting homology with the 36-bp VL30 direct repeats (underlining, Fig. 5B). The corresponding VL30 sequence contains two substitutions and three additional nucleotides relative to the MoMSV core sequence. Although the function(s) of these core nucleotides is not known, Laimins et al (25) have suggested that they play a critical role in the function of the enhancer element. Whether the 36-bp VL30 direct repeats have enhancer activity remains to be experimentally determined. In the orientation suggested by the PBS sequences, the enhancer homology is located 65 to 99 bp upstream from the TATATAA sequence.

The major structural features and putative transcriptional control signals of the integrated VL30 sequence are summarized diagrammatically in Fig. 7.

Whereas no polypeptides are known to be encoded by the LTRs of murine type C viruses, MMTV (a type B virus) encodes within its extended U3 region a hypothetical polypeptide of 198 amino acids (9) which has been translated in vitro (8). To date, there are no known VL30 polypeptides. A reading frame analysis of the LTR sequence, however, revealed an AUG codon at position 400 followed by an open reading frame capable of encoding a 57-amino acid polypeptide. This sequence terminates 31 bp before the end of the LTR (Fig. 3). Since this reading frame is relatively short, the possibility that it has occurred by chance cannot be discounted. Nevertheless, highly biologically active peptides of this size (EGF, for example) are common. In the orientation suggested in Fig. 7, the open reading frame lies 31 bp downstream from the putative transcription start site.

The findings that the major VL30-related RNA species induced after stimulation of quiescent cells with EGF and insulin exhibited an apparent chain length closely approximating an integrated VL30 gene and contained sequences derived from both LTR and non-LTR regions reinforce previous suggestions that at least some VL30 genes are specifically regulated by peptide growth factors (11). The present study differed from our previous study in that cells were maintained in serum-free media, rather than in 0.5% fetal bovine serum-containing media, for 48 h before stimulation. The absence of detectable VL30 RNA under these more defined conditions (Fig. 6A) suggests a nearly absolute dependence on peptide growth factor binding for VL30 gene expression. Although the molecular basis for this is unknown, it is tempting to speculate that LTR-specific sequences may play a role. It has been previously shown that specific sequences present in the MMTV LTR are required for glucocorticoid regulation of MMTV RNA synthesis (18, 26), and a similar mechanism could function in the present case. Experiments designed to test this possibility are in progress.

Several broad considerations arise from these studies. Although VL30 gene structure appears formally analogous to that displayed by certain classes of mobile genetic elements, at least some of these genes exhibit unique regulatory properties which clearly distinguish them from other known classes of transposon or retrovirus-like sequences in the mouse genome. To our knowl-

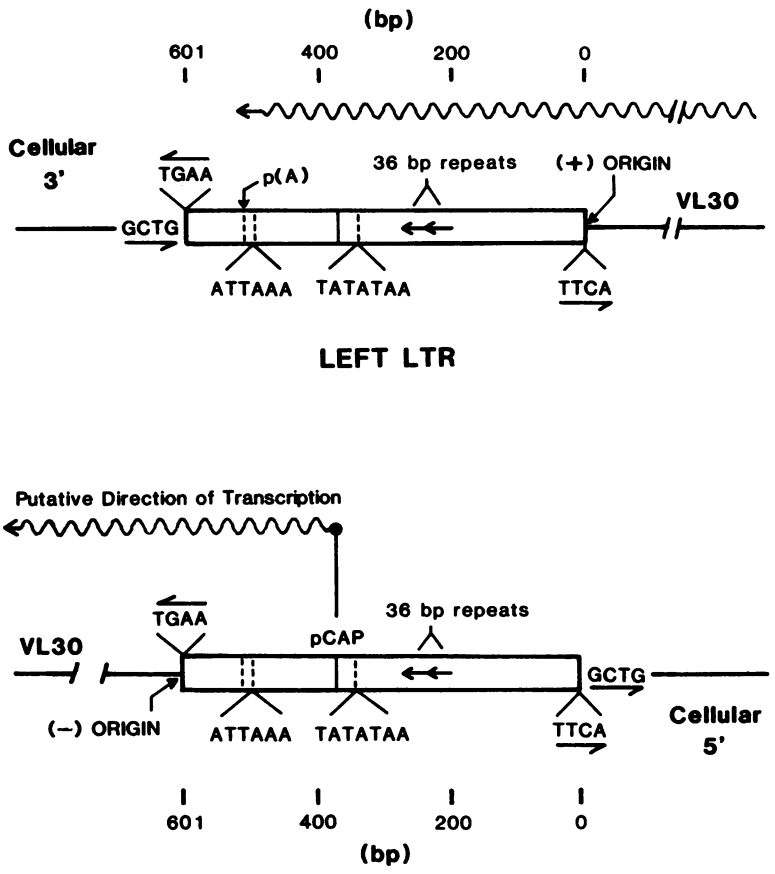

FIG. 7. General organization of the LTR regions of the VL30 element in clone BVL-1. Left and right refer to the orientation of the VL30 element as shown in Fig. 1. The positions of various structural features and possible control signals are indicated. (+) and (−) origins refer to regions bearing sequence homology to MuLV origins of DNA synthesis as shown in Fig. 4. pCAP and p(A) refer to putative CAP and polyadenylation sites, respectively. The 36-bp repeats are those showing partial homology to the MoMSV enhancer sequence as shown in Fig. 5B. The putative direction of transcription was assigned with reference to the minus and plus origin regions as discussed in the text.

edge, no other similar class of genes is specifically regulated by defined peptide growth factors. Although the in vivo role of these potent cellular mitogens remains uncertain, it seems likely that they function in regulating cell proliferation or cell differentiation or both. The elucidation of their mechanism of action may, therefore, be of fundamental importance in understanding how these processes are controlled in an orderly fashion during growth and development. Although many questions remain to be answered, the present data suggest that VL30 genes may prove useful in understanding how peptide growth hormones effect the differential expression of specific genes.

ACKNOWLEDGMENTS

We are grateful to Lucy Schmidt for excellent cell culture work, to G. Shipley for developing the serum-free culture conditions used in these studies, and to Patricia Hart for assistance in preparation of the manuscript. In addition, we are grateful to C. Liarakos for many helpful discussions.

This work was supported by Public Health Service grant GM 25510 from the National Institutes of Health and by the Mayo Foundation.

ADDENDUM

Since the submission of this manuscript, we became aware of an in press paper describing the nucleotide sequence of another VL30 LTR unit (A. Itin and E. Keshet, J. Virol. 47:656–659, 1983). Although considerable sequence differences were noted, the major structural features including the presence of inverted

terminal repeats, a tetranucleotide direct repeat of flanking cellular DNA, and LTR adjacent sequences bearing homology with MuLV origins of proviral DNA synthesis were conserved.

LITERATURE CITED

1. **Anderson, G. R., and K. C. Robbins.** 1976. Rat sequences of the Kirsten and Harvey murine sarcoma virus genomes: nature, origin, and expression in rat tumor RNA. J. Virol. **17**:335–351.
2. **Besmer, P., U. Olshevsky, D. Baltimore, D. Dolberg, and H. Fan.** 1979. Virus-like 30S RNA in mouse cells. J. Virol. **29**:1168–1176.
3. **Breathnach, R., and P. Chambon.** 1981. Organization and expression of eucaryotic split genes coding for protein. Annu. Rev. Biochem. **50**:349–383.
4. **Brutlag, D. L., J. Clayton, P. Freidland, and L. H. Kedes.** 1982. SEQ: a nucleotide sequence analysis and recombination system. Nucleic Acids Res. **10**:279–294.
5. **Courtney, M. G., P. K. Elder, D. L. Steffen, and M. J. Getz.** 1982. Evidence for an early evolutionary origin and locus polymorphism of mouse VL30 DNA sequences. J. Virol. **43**:511–518.
6. **Courtney, M. G., L. J. Schmidt, and M. J. Getz.** 1982. Organization and expression of endogenous virus-like (VL30) DNA sequences in nontransformed and chemically transformed mouse embryo cells in culture. Cancer Res. **42**:569–576.
7. **Dhar, R. W., 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.
8. **Dickson, C., R. Smith, and G. Peters.** 1981. *In vitro* synthesis of polypeptides encoded by the long terminal repeat region of mouse mammary tumour virus DNA. Nature (London) **291**:511–513.
9. **Donehower, L. A., A. L. Huang, and G. L. Hager.** 1981. Regulatory and coding potential of the mouse mammary tumor virus long terminal redundancy. J. Virol. **37**:226–238.
10. **Efstratiadis, A., F. C. Kafatos, and T. Maniatis.** 1977. The primary structure of rabbit beta globin mRNA as determined from cloned DNA. Cell **10**:571–585.
11. **Foster, D. N., L. J. Schmidt, C. P. Hodgson, H. L. Moses, and M. J. Getz.** 1982. Polyadenylated RNA complementary to a mouse retrovirus-like multigene family is rapidly and specifically induced by epidermal growth factor stimulation of quiescent cells. Proc. Natl. Acad. Sci. U.S.A. **79**:7317–7321.
12. **Garel, J. P., and G. Keith.** 1977. Nucleotide sequence of *Bombyx mori* L. $tRNA^{Gly}$. Nature (London) **269**:350–352.
13. **Getz, M. J., P. K. Elder, E. W. Benz, Jr., R. E. Stephens, and H. L. Moses.** 1976. Effect of cell proliferation on levels and diversity of poly(A)-containing mRNA. Cell **7**:255–265.
14. **Giri, C. P., C. P. Hodgson, P. K. Elder, M. G. Courtney, and M. J. Getz.** 1983. Discrete regions of sequence homology between cloned rodent VL30 genetic elements and AKV-related MuLV provirus genomes. Nucleic Acids Res. **11**:305–319.
15. **Hagenbuchle, O., R. Bovey, and R. A. Young.** 1980. Tissue-specific expression of mouse α-amylase genes: nucleotide sequence of isoenzyme mRNAs from pancreas and salivary gland. Cell **21**:179–187.
16. **Hobbs, A. A., and J. M. Rosen.** 1982. Sequence of rat α- and γ-casein mRNAs: evolutionary comparison of the calcium-dependent rat casein multigene family. Nucleic Acids Res. **10**:8079–8098.
17. **Howk, R. S., D. H. Troxler, D. Lowy, P. H. Duesberg, and E. M. Scolnick.** 1978. Identification of a 30S RNA with properties of a defective type C virus in murine cells. J. Virol. **25**:115–123.
18. **Huang, A. L., M. C. Ostrowski, D. Berard, and G. L. Hager.** 1981. Glucocorticoid regulation of Ha-MuSV p21 gene conferred by sequences from mouse mammary tumor virus. Cell **27**:245–255.
19. **Keshet, E., and A. Itin.** 1982. Patterns of genomic distribution and sequence heterogeneity of a murine "retrovirus-like" multigene family. J. Virol. **43**:50–58.
20. **Keshet, E., and Y. Shaul.** 1981. Terminal direct repeats in a retrovirus-like repeated mouse gene family. Nature (London) **289**:83–85.
21. **Keshet, E., Y. Shaul, J. Kaminchik, and H. Aviv.** 1980. Heterogeneity of "virus-like" genes encoding retrovirus-associated 30S RNA and their organization within the mouse genome. Cell **20**:431–439.
22. **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.
23. **Korn, L. J., and D. D. Brown.** 1978. Nucleotide sequence of *Xenopus borealis* oocyte 5s DNA: comparison of sequences that flank several related eukaryotic genes. Cell **15**:1145–1156.
24. **Lai, E. C., J. P. Stein, J. F. Catterall, S. L. C. Woo, M. L. Mace, A. R. Means, and B. W. O'Malley.** 1979. Molecular structure and flanking sequences of the natural chicken ovomucoid gene. Cell **18**:829–842.
25. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **79**:6453–6457.
26. **Lee, F., R. Mulligan, P. Berg, and G. Ringold.** 1981. Glucocorticoids regulate expression of dihydrofolate reductase cDNA in mouse mammary tumor virus chimeric plasmids. Nature (London) **294**:228–232.
27. **Levinson, B., G. Khoury, G. Vande Woude, and P. Gruss.** 1982. Activation of SV40 genome by 72-base pair tandem repeats of Moloney sarcoma virus. Nature (London) **295**:568–572.
28. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning, a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
29. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
30. **McReynolds, L., B. W. O'Malley, A. D. Nisbet, J. E. Fothergill, D. Givol, S. Fields, M. Robertson, and G. G. Brownlee.** 1978. Sequence of chicken ovalbumin mRNA. Nature (London) **273**:723–728.
31. **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.
32. **Proudfoot, N. J.** 1977. Complete 3′ noncoding region sequences of rabbit and human beta globin messenger RNAs. Cell **10**:559–570.
33. **Scolnick, E. M., W. C. Vass, R. S. Howk, and P. H. Duesberg.** 1979. Defective retrovirus-like 30S RNA species of rat and mouse cells are infectious if packaged by type C helper virus. J. Virol. **29**:964–972.
34. **Shapiro, J.** 1979. Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. U.S.A. **76**:1933–1937.
35. **Sherwin, S. A., U. R. Rapp, R. E. Benveniste, A. Sen, and G. J. Todaro.** 1978. Rescue of endogenous 30S retroviral sequences from mouse cells by baboon type C virus. J. Virol. **26**:257–264.
36. **Tu, C. P. D., and S. N. Cohen.** 1980. 3′ end labeling of DNA with [α-^{32}P] cordycepin-5′-triphosphate. Gene **10**:177–183.
37. **Van Beveren, C., J. G. Goddard, A. Berns, and I. M. Verma.** 1980. Structure of Moloney murine leukemia viral DNA: nucleotide sequence of the 5′ long terminal repeat and adjacent cellular sequences. Proc. Natl. Acad. Sci.

U.S.A. 77:3307–3311.
38. **Van Beveren, C., E. Rands, S. K. Chattopadhyay, D. R. Lowy, and I. M. Verma.** 1982. Long terminal repeat of murine retroviral DNAs: sequence analysis, host-proviral junctions and preintegration site. J. Virol. 41:542–556.
39. **Varmus, H. E.** 1982. Form and function of retroviral proviruses. Science 216:812–820.
40. **Zarucki-Schulz, T., S. Y. Tsai, K. Itakura, X. Soberon, R. B. Wallace, M. J. Tsai, S. L. C. Woo, and B. W. O'Malley.** 1982. Point mutagenesis of the ovalbumin gene promoter sequence and its effect on *in vitro* transcription. J. Biol. Chem. 257:11070–11077.