A novel sequence segment and other nucleotide structural features in the long terminal repeat of a BALB/c mouse genomic leukemia virus-related DNA clone

Chin-Yih Ou, Lawrence R.Boone+ and Wen K.Yang*

University of Tennessee-Oak Ridge Graduate School of Biomedical Sciences, and Biology Division, Oak Ridge National Laboratory, P.O. Box Y, Oak Ridge, TN 37830, USA

Received 10 January 1983; Revised and Accepted 27 July 1983

ABSTRACT

A recombinant DNA clone, named AL10, that contains murine leukemia virus (MuLV) related sequences was isolated from BALB/c mouse chromosomal DNA and examined in detail. Restriction endonuclease mapping revealed that the 10.5 kbp EcoRI insert consists of a 3.6 kbp left flanking cellular DNA region and a 6.9 kbp MuLV-related region that has a typical proviral LTR-gag-pol-env structure up to the EcoRI site in the env gene region. Comparison of the AL10 map with ecotropic and xenotropic virus isolates revealed many common restriction sites in the LTR and pol gene regions, but much fewer in the leader and gag regions. A stretch of 1,700 nucleotides containing the cell-provirus junctional region was sequenced and revealed transcriptional consensus signals and other structural features characteristic of MuLV LTRs, as well as two distinctive features: (a) a sequence of ~170 bp with direct and inverted terminal repeats not seen in infectious MuLV LTRs was identified in the U3 region between the "enhancer" region and the "CAT" box. This novel segment or its homologous sequences appear to be present in most of the endogenous MuLV-related LTRs and in other chromosomal locations of the mouse (b) The tRNA primer binding site is not complementary to proline tRNA, the primer for all known MuLVs, but is a 17/18 match with rat glutamine tRNA. The integration site of AL10 provirus was in a unique DNA region but contained an "Alu"-like short interdispersed repeat in the 5' adjacent cellular region. The AL10 proviral integration found in BALB/c was also apparent in RFM, AKR and SENCAR mouse cells but not in cells of NFS/N, C3H, HRS/J, SC-1, and a California Lake Casitas wild mouse.

INTRODUCTION

It is now well recognized that retroviral long terminal repeats (LTR), formed from 5' and 3' terminal sequences of viral genomic RNA by reverse transcription and located at the ends of both the linear DNA intermediate and integrated proviruses, are important functional units for gene expression and presumably for gene integration (reviews in 1-4). The LTRs of murine leukemia viruses (MuLV) and sarcoma viruses (MuSV) have been shown to contain consensus sequences of the eukaryotic transcription unit as well as structural features similar to transposable gene elements (5-9), to precisely direct RNA synthesis in an in vitro transcription assay (10-11) and to

markedly promote and/or enhance oncogene expression when linked to a cellular oncogene (12, 13). While all these results were obtained with horizontally transmissible retroviruses, very little is known in this regard about the MuLV-related sequences present in normal mouse chromosomal DNA. There are as many as 20-50 copy equivalents of these genes per haploid genome of the mouse (14-19). Genetic and molecular analyses have indicated that ecotropic MuLVs (infectious for mouse cells) and xenotropic MuLVs (restricted in mouse cells) are induced, separately, from a few distinct chromosomal loci of the mouse (20-24). The majority of endogenous MuLV related sequences have not been shown to be expressed as infectious virus.

We have taken the molecular cloning approach to study the organization of endogenous MuLV genomes of the BALB/c mouse. This strain contains a N-ecotropic MuLV locus on chromosome 5 (21),a xenotropic MuLV locus on chromosome 1 (23), and sequences for the generation of a B-tropic MuLV (25, 26) in addition to the abundance of uncharacterized sequences.

We have isolated 14 recombinant DNA clones containing MuLV-related sequences from BALB/c mouse genomic DNA. One of the clones, AL10, has been analyzed in detail; it has a typical proviral structure of cell-LTR-gag-pol-env (up to the EcoRI site in the env region), which is apparently distinct from those of ecotropic and xenotropic MuLVs of BALB/c mice and contains an LTR longer than all known infectious MuLV LTRs. We determined the nucleotide sequences of the AL10 LTR and its adjacent regions to search for possible features related to the control of this MuLV-related sequence in the cell.

MATERIALS AND METHODS

Cells, Viruses and DNA Clones. Inbred BALB/c, C3H, AKR, and RFM/Un mice were purchased from Cumberland View Farms, Clinton, Tennessee. Inbred Swiss NFS/N mice, Lake Casitas wild mouse cells and a human fibrosaroma HT 1080 cell line were initially supplied by Dr. R. H. Bassin (NIH), Dr. M. B. Gardner (USC) and Dr. R. M. McAlister (UCLA), respectively. Two ecotropic MuLV strains (WN1802N and Gross) and SC-1 cells used for virus propagation were originally obtained from Drs. Hartley and Rowe of NIH. Genomes of these ecotropic MuLVs and RFV, an ecotropic endogenous MuLV isolated from RFM/Un mice have been molecularly cloned in lambda phage and plasmid pBR322 vectors in this laboratory (27, 28).

Cellular DNA Preparation. Fresh BALB/c mouse livers were used to isolate chromosomal DNA, according to the standard procedure (8). To isolate DNA from cultured BALB/c mouse embryo cells, extraction procedures of Hirt

(29) were employed first. The chromatin pellet was dissolved in a solution of 10 mM Tris Cl (pH 7.6) and 10 mM EDTA, digested with pronase and the DNA isolated by phenol chloroform extraction.

Nick-translated DNA probes. The molecular cloning and characterization of pG100 (Gross strain of MuLV) and pWN41 (ecotropic WN1802N virus of BALB/c mouse) have been described (27, 28). Both MuLV DNAs contained two tandemly-repeated LTRs and were cloned into pBR322 through the single HindIII sites in their pol genes. Several restriction fragments derived from these two clones were used as molecular probes. SmaI/SmaI LTR fragments containing the entire LTR sequences from both clones and PstI/KpnI fragment of pWN41 containing a portion of LTR sequences were used as complete LTR probes and a partial LTR probe, respectively. These DNA fragments were isolated from a 5% polyacrylamide gel by the diffusion-extraction procedure (30). A 1.1 kbp HindIII/SmaI fragment of pG100 was isolated from an 0.7% agarose gel by the electroelution procedure (31) and was used as a pol-env probe. Gag specific and flanking region BB and BA probes were derived from a 1.4 kbp BglII/BglII, a 1.2 Kbp BamHI/BglII and a 1.1 Kbp BamHI/AvaI fragment of AL10, respectively. An eco specific probe specific for the env region of ecotropic MuLVs was derived from a 300 bp SmaI/SmaI fragment of pRFV. All the DNA fragments were labeled with [^{32}P] by the nick-translation procedure (32).

Digestion of DNA with Restriction Endonucleases. Restriction enzymes were purchased from Bethesda Research Laboratories and New England Biolabs, and digestion of DNA was carried out according to the specifications recommended by the manufacturers. For determinations of sequential restriction enzme sites in the cloned DNA, the partial digestion technique of Smith and Birnstiel (33) was employed.

Gel Electrophoresis and Blotting of DNA. Analysis of DNA was carried out essentially according to the established methods of horizontal submerged agarose gel electrophoresis/transfer to nitrocellulose membrane or diazobenzyloxy methyl-paper/molecular hybridization autoradiography, as previously described (34, 37).

Molecular Cloning. High molecular weight chromosomal DNA was digested to completion with EcoRI and the fragments separated by electrophoresis in a 0.6% gel of low-temperature-melting agarose (Seakem). The gel was divided into fractions and the DNA recovered by melting and phenol extraction.

DNA from a 10-20 Kbp fraction was cloned into Charon 9 by the protocol of Blattner (38) and approximately 55,000 recombinant phages were screened by the Benton and Davis (39) procedure using ^{32}P labeled cDNA of WN1802B MuLV

RNA initially (34), and confirmed with ^{32}P labeled nick-translated pG100 DNA.

The 10.5 kbp insert of the AL10 clone was transferred into pBR322 and maintained in E. coli C600 according to standard procedures (40). An internal BamHI fragment of the AL10 insert was subcloned into the BamHI site of pBR322 and designated as pAL10B3.

Subcloning of a Sau3A/Sau3A DNA fragment, contained in the novel sequence segment found in the AL10 LTR, was carried out as follows: pAL10B3 DNA was digested with BglII. A 718 bp fragment containing most of the U3 region of LTR was isolated from a 5% polyacrylamide gel and then further digested with Sau3A enzyme. A 155 bp Sau3A/Sau3A fragment representing the novel sequence was isolated from a second 5% polyacrylamide gel and inserted into the BamHI site of pBR322. This clone was designated pAL10-170. A 714 bp EcoRI/SalI fragment of pAL10-170 containing the inserted Sau3A/Sau3A fragment and part of the adjacent pBR322 DNA was isolated from a 5% polyacrylamide gel and nick-translated.

Nucleotide Sequencing. The detailed procedures described by Maxam and Gilbert (41) were followed. Single end-labeled DNA fragments were isolated either by the strand separation method or by the additional restriction enzyme digestion method, depending on feasibility.

RESULTS

Fourteen Charon 9 recombinant DNA clones, designated AL1 through AL14, were isolated from a 10-20 kbp fraction of EcoRI-digested BALB/c mouse DNA by hybridization to the nick-translated pG100 DNA (Gross Passage A MuLV) as described in materials and methods. All clones contained sequences which hybridized to a LTR probe and a pol-env probe derived from pG100. However, none hybridized to an ecotropic MuLV-specific env probe derived from pRFV (data not shown). The sizes of the mouse insert DNAs in clone AL2, 3, 10, 11, 12, 13 and 14 are 20.5, 20.5, 10.5, 19.5, 19.5, 20, and 13.5 kbp respectively. All clones also contained sequences homologous to a gag-specific probe derived from the AL10 clone (see below). The AL10 clone was chosen for detailed analysis.

Restriction Enzyme Mapping of AL10 Clone. The 10.5 kbp insert of AL10 clone is larger than the estimated 8.8 kbp MuLV genome and would be expected to contain flanking cellular sequences. To determine the physical arrangement of proviral and cellular DNA sequences in this clone, we isolated the insert and examined its restriction enzyme-digested fragments with the nick-translated probes of three DNA sequences: (a) the insert itself to identify

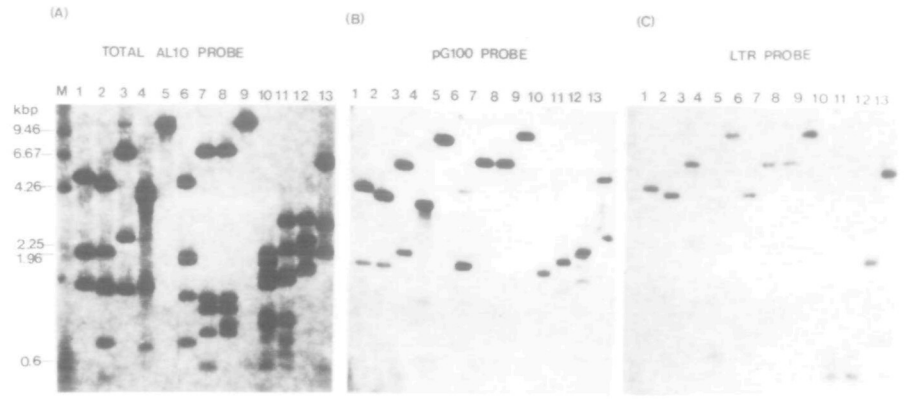

FIG. 1. Representative autoradiograms for restriction enzyme/gel blot analysis of the AL10 insert DNA with the use of nick-translated probes of total AL10 insert (panel A), pG100 MuLV DNA (panel B) and LTR of pG100 (panel C). The AL10 insert, isolated by EcoRI digestion and electrophoretic gel elution, was digested by XhoI (lane 1), XhoI plus SalI (lane 2), SalI (lane 3), SmaI (lane 4), PvuI (lane 5), KpnI (lane 6), PstI plus HindIII (lane 7), PstI (lane 8), HindIII (lane 9), BglI plus BglII (lane 10), BglII (lane 11), BglI (lane 12), and BamHI (lane 13). The digested DNAs were separated in a 0.7% agarose gel (2 volts/cm, 14 hours) and transferred to diazobenzyloxymethyl-papers according to the Southern procedure (48). The same papers containing the transferred DNA fragments were examined successively for hybridization with the three molecular probes, employing an alkaline de-probing step prior to the re-hybridization procedures.

all insert fragments; (b) the total MuLV sequence of pG100 to identify fragments containing MuLV-related sequences; (c) a SmaI/SmaI cut fragment from the permuted LTRs of pG100. Representative gel blot patterns of DNA fragments produced by single and double restriction enzyme digestion are presented in Figure 1. All of the pG100-related sequences could be placed in a single domain consisting of the right two-thirds of the 10.5 kbp DNA insert, whereas all sequences not hybridized to the pG100 probe (except a 0.2 kbp SmaI/EcoRI fragment) were located in the left one-third of the insert. The right-end location of the 0.2 kbp SmaI/EcoRI fragment suggests that it represents an AL10-specific env sequence sharing no homology with *Gross MuLV. All fragments containing the LTR sequence can be placed at the* junction of the two domains. The insert was cut out by EcoRI digestion, end-labeled with [^{32}P] and then cut with HindIII to generate a 10.1 kbp single end-labeled fragment representing most of the AL10 insert sequences except the left-end 0.4 kbp region. The precise restriction sites for 8 restriction enzymes in the 10.1 kbp fragment were determined by means of

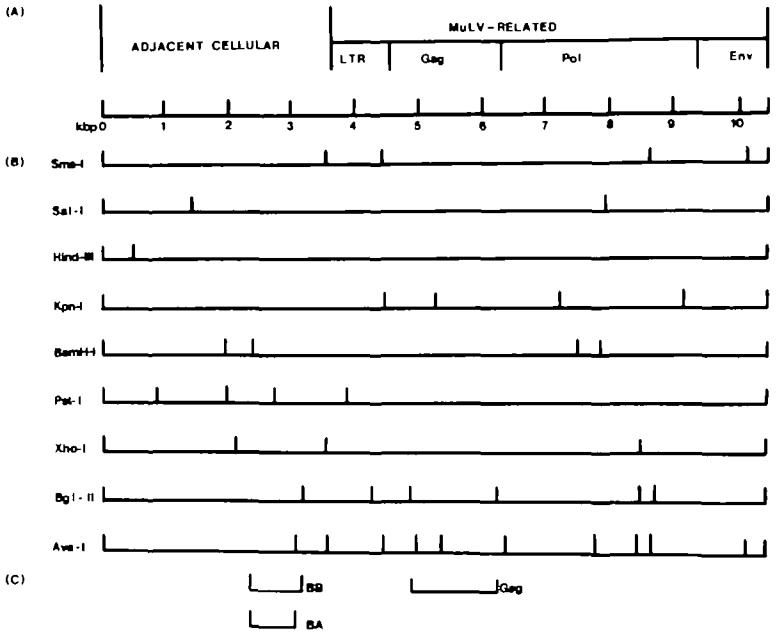

FIG. 2. Restriction enzyme map of AL10 DNA. The corresponding regions of the MuLV-related and adjacent cellular sequences are indicated in the top drawing (A). The three DNA fragments used as specific probes are indicated in (C).

the partial enzyme digestion procedure of Smith and Birnstiel (33). With all the results combined, a physical map of the AL10 insert was constructed (Fig. 2). The 10.5 kbp AL10 insert is composed of a 6.9 kbp MuLV-related proviral sequence and 3.6 kbp of upstream flanking cellular sequence. The EcoRI site in the env gene region prevented the isolation of complete provirus copy with the remaining 2 kbp sequence. The AL10 LTR is approximately 700 bp in length, which is about 100 bp longer than the Moloney MuLV LTR (which contains a 75 bp repeat) and about 170 bp longer than the LTR of WN1802N and RFV, which are the endogenous N-tropic ecotropic MuLVs of BALB/c (27) and RFM/Un mice (28), respectively.

Nucleotide Sequence Analysis. We subcloned a BamHI internal fragment (position 2.4-7.5 kbp, Fig. 2) of AL10 into the BamHI site of pBR322. The DNA of this clone, pAL10B3, was used for nucleotide sequence analyses by the Maxam-Gilbert procedure (30). A sequence of 1,700 nucleotides was obtained, which encompasses 500 nucleotides of upstream non-viral cellular

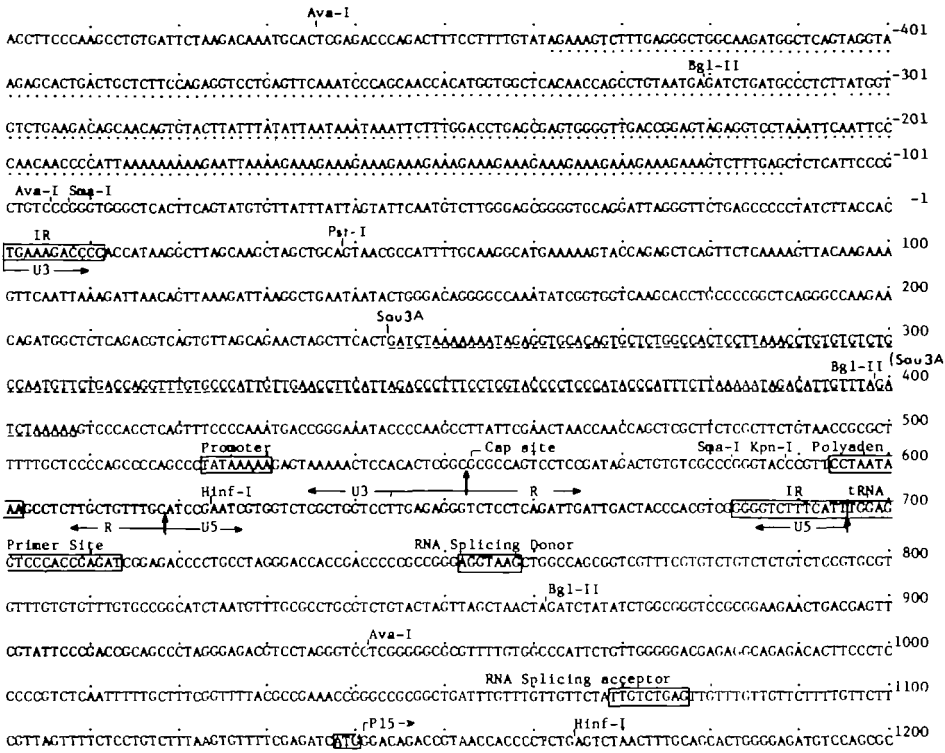

FIG. 3. Nucleotide sequence of upstream cellular (-1 to -500), LTR (1-695), leader (696-1137) and part of p15 coding sequence of gag gene in the AL10 clone. Only the positive strand corresponding to retroviral genomic RNA sequences is shown. Several landmarks are indicated and the ATG initiation codon of the gag gene is boxed (1138-1140).

region, the 5' LTR, the leader and the beginning of the gag protein coding region (Fig. 3).

Landmarks and Transcriptional Signals. Both 5' and 3' LTRs of the unintegrated linear DNA intermediate of known MuLVs are bound by 13 bp inverted repeats, AATGAAAGACCCC/GGGGTCTTTCATT (5, 6). The integrated form of MuLV DNA, while colinear with the unintegrated linear DNA, lacks the terminal nucleotides, AA/TT on both LTRs, apparently due to the integration process (8, 9). In the 1700 nucleotide sequence, the identical inverted repeats were located at positions 1-11 and 683-693, but without the terminal AA/TT. This indicates that the AL10 LTR is 695 bp long and that this vertically inherited proviral DNA of BALB/c mice may have arisen originally by an integration process similar to that of exogenous murine retroviral DNAs (8, 9).

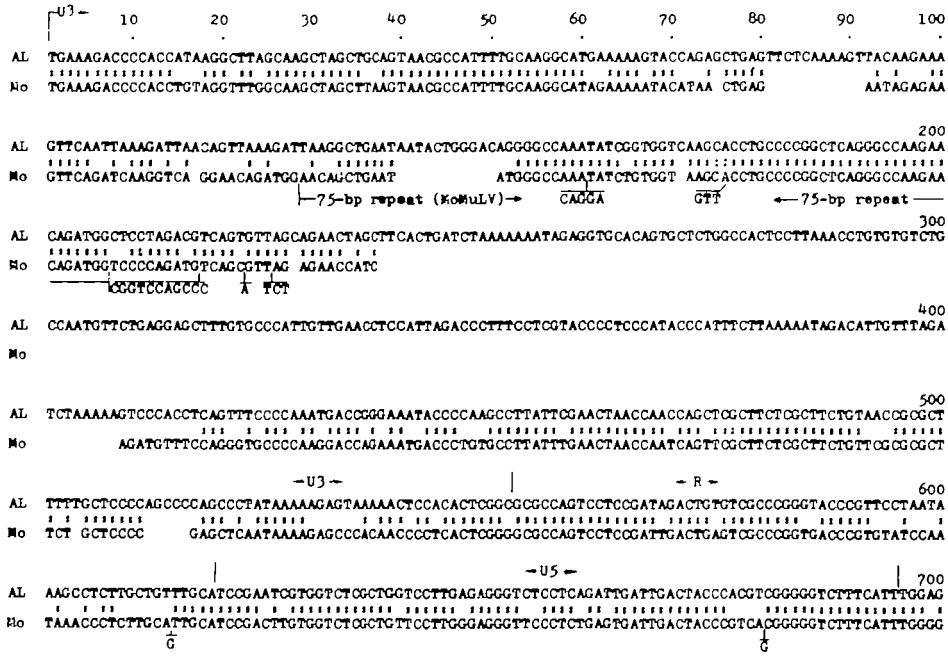

FIG. 4. Comparison of the LTR sequences of the AL10 clone and the Moloney MuLV. The Moloney MuLV sequence (5, 6, 8) is matched to the AL10 sequence at homologous regions, hence showing deletions or insertions.

The retroviral LTR structure is generally presented as U3-R-U5, to indicate that the U3-R and the R-U5 correspond, respectively, to defined sequences at the 3' and the 5' ends of viral genomic RNA (42). Comparison of the nucleotide sequences of the Moloney MuLV LTR at these boundaries suggests that the AL10 LTR is comprised of a 552 bp U3, a 66 bp R and a 77 bp U5. These boundaries correspond to positions relative to the following transcriptional signals: (i) the presumed RNA polymerase II binding site or the Hogness-Goldberg box, TATAAAAA, positions 523-530, which is 23 to 30 bp upstream from the RNA "cap" site (U3/R junction); (ii) the presumed polyadenylation signal, CCTTAATAAA, at positions 594-602, which is 17 to 25 bp upstream from the poly A acceptor site (R/U5 junction); and (iii) the presumed CAT box, CCAAC, at positions 466-470, which is 83 to 87 bp upstream from the cap site.

A Novel Sequence Segment in the U3 of AL10 LTR. A comparison of the AL10 LTR and the published sequences for Moloney MuLV and MuSV (5, 6, 8) was made (Fig. 4). The U5 and R regions showed considerable homology (77%),

whereas the U3 region is marked by alternating homologous and nonhomologous segments, with major differences present in the middle portion, i.e., between positions 80 and 418 of the AL10 LTR. The 75 bp "enhancer" sequence (42), which is tandemly repeated in Moloney MuLV, is not duplicated in the AL10 LTR and shows alterations at the 5' side. The most marked difference found in the AL10 LTR is a nonhomologous 170 bp segment, which contains, near its ends, 10 bp direct repeats (GATCTAAAAA at positions 244-253 and 399-408), 9 bp direct repeats (AAAAATAGA at positions 251-259 and 381-389) and 8 bp inverted repeats (GATCTAAA at positions 244-251 and TTTAGATC at positions 395-402). No sequence homology was detected between this 170 bp segment and other regions of the AL10 LTR, indicating that it was not derived by sequence duplication or rearrangement.

Occurrence of the Novel Sequence of AL10 LTR in other MuLV-related Clones and in Chromosomal DNA of the BALB/c Mouse. The LTR of AL10 is similar to the LTR of many MuLV proviruses (9, 27, 28, 43, 44) in possessing a PstI restriction site, 38 bp from the left end (positions 34-39), and a KpnI restriction site, 117 bp from the right end (positions 584-589); between these two sites the novel 170 bp segment is located. To determine whether or not the LTR of other MuLV-related clones that we isolated from BALB/c mice have the similar sequence property, the DNA of these clones as well as AL10 and BALB/c mouse-derived N-tropic MuLV pWN41 clone (27) were cleaved by combined PstI-KpnI digestion and analyzed by gel blot hybridization with a representative total LTR probe, prepared from pWN41, and a specific probe prepared from the AL10 novel segment. As shown in Fig. 5B, in the gel lane of every DNA sample the LTR probe revealed a single strongly-hybridizing band in the 600 bp range and one large weakly-hybridizing band. The strongly-hybridizing band varied in three sizes and hence designated as band a, band b and band c. It is known from the nucleotide sequences (Fig. 2 and our unpublished data of pWN41 LTR) that the band b of AL10 and the band c of pWN41 represent the PstI/KpnI internal LTR fragments and are of 540 bp and 372 bp, respectively, and that the weakly-hybridizing bands represent the right-end 117 bp LTR sequence linked to other viral sequences. The band a, an apparent 600 bp PstI/KpnI internal LTR fragment, was found in all other MuLV-related clones, except AL14 which was similar to AL10 in containing band b. Fig. 5A shows that both band a and band b but not band c hybridized strongly to the specific probe of the AL10 novel segment. Thus, the results indicated that all 14 BALB/c mouse DNA-derived MuLV-related clones contained the PstI and KpnI sites as well as 170 bp novel segment or homologous sequences between

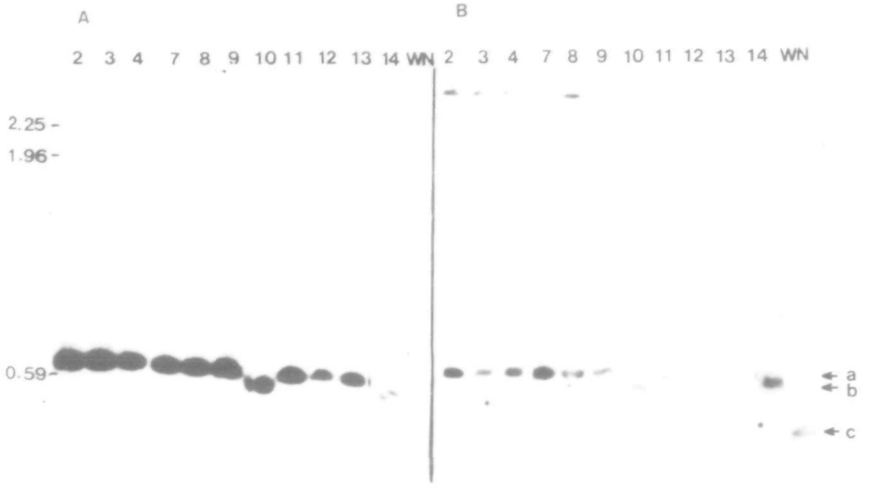

FIG. 5. Southern gel blot analyses for the presence of the AL10 170 bp segment or its homologous sequences in the LTRs of MuLV-related proviral sequences. All the DNAs examined are Charon 9 lambda phage recombinants (clones AL2, 3, 4, 7, 8, 9, 10, 11, 12, 13 and 14). WN (a WN41 clone) designates the WN1802N-tropic MuLV isolate of the BALB/c mouse (27). PstI/KpnI digested DNAs of the clones were separated by horizontal electrophoresis in 1.5% agarose gel (3 volts/cm, 16 hr) and transferred to nitrocellulose membrane. (A), EcoRI/ SalI fragment of pAL10-170 subclone (molecular cloning, Materials and Methods) and (B), a 527 bp SmaI/SmaI bp fragment containing the complete WN1802N (WN41) viral LTR were used to prepare nick-translated probes for hybridization with the membranes. Note that each MuLV-related endogenous sequence clone gave two bands with the LTR probe (the PstI/KpnI fragment and a larger one containing the 3' 117 bp portion of LTR). The WN clone which was originally isolated from closed circular viral DNA containing two LTRs in tandem generated 3 bands; the third band, which was the 155 bp Pst/Kpn LTR fragment ran out of bound and was not shown. Weak hybridization signals (1 Kbp) identified by AL10-170 probe in some clones were not characterized.

these two restriction sites in their LTR (data demonstrating band a in clones AL1, AL5, and AL6 by the LTR probe and the novel segment probe are not shown).

To determine the significance of the above-mentioned findings, we made similar PstI/KpnI-gel blot analyses with the chromosomal DNA of BALB/c mice from which the MuLV-related clones were derived. Results, shown in Fig. 6, indicate that the BALB/c chromosomal DNA contains all three PstI/KpnI internal LTR fragments (band a/band b/ band c in the 600 bp range) and that the band a and band b types constitute the major population. The total pWN41 LTR probe detected also a large number of high molecular weight DNA fragments (lane 2, Fig. 6) most of which were considered to contain the 5' (38 bp) or

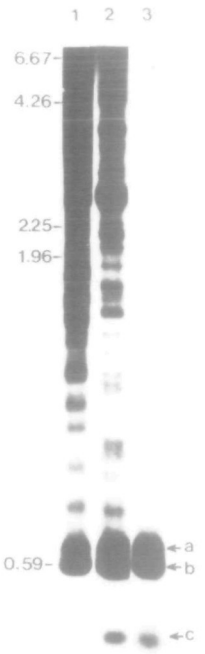

FIG. 6. Southern gel analyses revealing the presence of homologous sequences to the AL10-170 segment in endogenous MuLV-related LTRs and in other chromosomal locations of the BALB/c mouse. Ten micrograms of BALB/c embryonic DNA was digested with PstI and KpnI enzymes and analyzed by Southern blotting procedures as described in Fig. 4. Lane 1, AL10-170 probe; lane 2, complete LTR probe (SmaI/SmaI fragment of pWN41); lane 3, partial LTR probe (PstI/KpnI fragment of pWN41). Specific activity of these probes was 2-5 × 10^8 cpm/μg DNA. Approximately 0.5 × 10^6 cpm per lane was used. Autoradiogram was carried out at -80°C for 20 hours with the use of DuPont intensifying screens.

the 3' (117 bp) sequences of the LTR since these were not detected by a probe prepared from the 372 bp PstI/KpnI fragment of pWN41 LTR (lane 3, Fig. 6). As expected, the AL10 novel segment probe showed hybridization with the band a and the band b, but not the band c, of the chromosomal DNA (lane 1, Fig. 6). Quite strikingly, the novel segment probe also hybridized strongly with PstI/KpnI digest of BALB/c mouse chromosomal DNA in a highly multiple and slightly smearing fashion. It is thus apparent that the AL10 LTR ~170 bp novel segment or homologous sequences are repeated not only in most of the endogenous MuLV proviral LTRs but also in other chromosomal locations not related to the MuLV LTRs of pWN41 type.

Distinct tRNA Primer Binding Site. All infectious ecotropic MuLVs are known to utilize proline tRNA as the primer (45, 46). The tRNA binding sites of cloned MuLV DNAs, such as Moloney (5), AKR (9), WN1802N, WN1802B, Gross passage A and RFV (our unpublished data), have been found to be consistently TGGGGGCTCGTCCGGGAT, which is complementary to the 18 nucleotides at the 3' end of mammalian proline tRNA. The sequence of this site in the AL10 clone,

however, is TGGAGGTCCCACCGAGAT (Fig. 3). This sequence, with the dispersed 6 base changes, obviously cannot bind proline tRNA by base-pairing. The recently published sequence of rat liver major glutamine tRNA (47) is a 17/18 complementary match with the AL10 primer binding site.

The Adjacent 5' Cellular Region. The tetranucleotide, CCAC, immediately to the left of the 11 bp inverted repeat of the AL10 LTR, is not identical to any of the tetranucleotides reported to be an integration site of integrated MuLV DNAs (8, 9). This is consistent with the idea that the short cellular sequence duplicated at the site of retroviral DNA integration is not uniform (1).

It is apparent that the sequence located at positions -439 to -113 (underlined with dots) is related to the "Alu"-like repetitive sequence of rodents (48). Like most of the other vertebrate interdispersed repetitive sequences, this sequence has the common structure of [Direct repeat] ["Alu"] [A-rich] [Direct repeat]. In this case, the direct repeats are AGAAAGTCTTTGAG at positions -439 to -426 and -126 to -113. The main "Alu" structure appears to be a composite of two segments: a 175 bp type 2 "Alu-equivalent" (positions -425 to -251) and a 63 bp segment (positions -252 to -188). The 175 bp segment is a near perfect match with the "Alu-equivalent" sequence found in the intron of a rat growth hormone gene (49, 50) while the 63 bp segment appears not to match with any known Alu-like sequence. The A-rich sequence, $A_9GA_2TTA_3$ $(AGAA)_{11}$ at positions -187 to -127, and with multiple AGAA repeats, appears to be distinctly different from other Alu-sequence-associated A-rich segments.

Characterization of the Integration Site of the AL10 Provirus in BALB/c Mouse DNA. As shown in the physical map (Fig. 2), the flanking cellular region of the AL10 clone contains a BamHI site and a BglII site at 1.5 kbp and 0.3 kbp, respectively, upstream of the LTR. This 1.2 kbp BamHI/BglII fragment (designated as BB in Fig. 2C) was isolated and used for preparing a nick-translated probe. Gel blot analysis indicated that this probe hybridized extensively and uniformly throughout the gel lane of EcoRI-digested BALB/c mouse DNA while no hybridization with a human DNA preparation was observed (Fig. 7A). The extensive and uniform hybridization pattern, in contrast, disappeared when a BamHI/AvaI fragment (designated as BA in Fig. 2C) was used as the probe (Fig. 7B). These results confirm the sequence analysis that a mouse specific, short, interspersed and repetitive sequence is present at about 300 bp upstream from the 5' LTR of the AL10 provirus (Fig. 3).

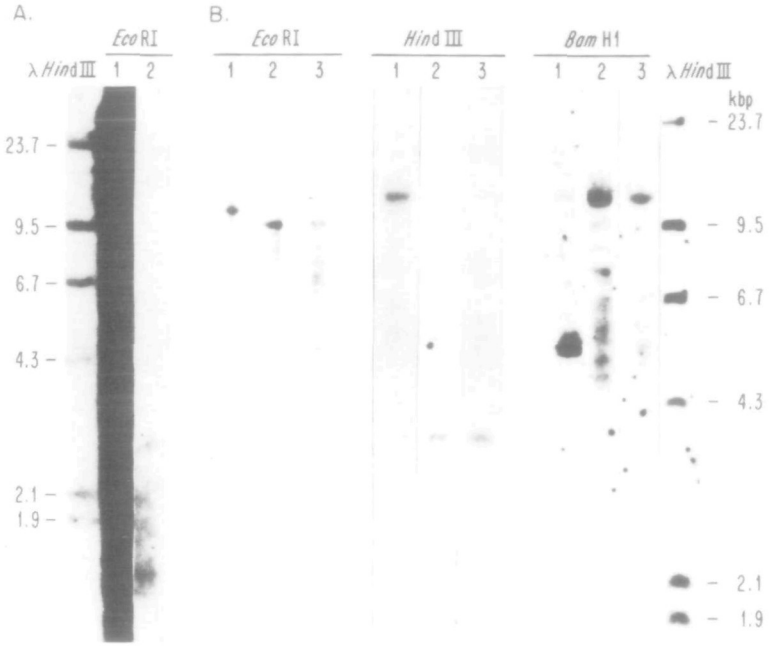

FIG. 7. Characterization of the AL10 provirus integration site by gel blot analysis with nick-translated probes made from the flanking cellular sequences of the AL10 DNA insert. (A) EcoRI-digested BALB/c strain mouse DNA (lane 1) and HT 1080 human fibrosarcoma cell DNA (lane 2), analyzed with the "BB" probe. (B) EcoRI-, HindIII-, and BamHI-digested BALB/c strain mouse DNA (lane 1), Swiss NFS/N strain mouse DNA (lane 2) and a California Lake Casitas mouse cell DNA (lane 3) analyzed with "BA" probe. HindIII fragments of lambda phage DNA (HindIII) served as the DNA size marker. See Fig. 2C for the location of "BB" and "BA" sequences.

With the use of the BA probe, a single 10.5 ± 0.2 kbp EcoRI-generated fragment, a single 12.8 ± 0.2 kbp HindIII-generated fragment and a single 5.3 ± 0.1 kbp BamHI-generated fragment were revealed in the BALB/c mouse DNA (Fig. 7B). The sizes of the BamHI and the EcoRI fragments are exactly as predicted from the physical map of the AL10 clone. The 12.8 kbp BA-hybridizing HindIII fragment of BALB/c mouse DNA may consist of a 3.2 kbp flanking sequence on the left, a presumed 9 kbp provirus and a 0.6 kbp flanking sequence on the right.

The BA probe was also employed to analyze the restriction endonuclease/gel blots of the DNA preparations from livers of inbred Swiss NFS/N strain mice and embryo cells of a California Lake Casitas wild mouse. These two mouse varieties were selected for study because of their origin from

different geographical locations in the world. As shown in Fig. 7B the BA-homologous sequence of both the NFS/N and the Lake Casitas mouse was detected in a single 9.9 ± 0.2 kbp EcoRI fragment, a single 3.7 ± 0.2 kbp HindIII fragment and a single 13.0 ± 0.5 kbp BamHI fragment, which were all distinctly different from those obtained from BALB/c mouse DNA. We also analyzed chromosomal DNA of 6 other inbred and wild mice with the BA probe. The EcoRI and HindIII fragment of AKR, RFM and SENCAR mice detected by BA probe are of the same size as their corresponding fragments found in BALB/c mice. On the other hand, EcoRI and HindIII DNA fragments of C3H, HRS/J mouse and SC-1 cells are the same as those found in NFS and Lake Casita wild mice (data not shown). The 3.7 kbp HindIII fragment of NFS/N, Lake Casitas, C3H, HRS/J mice and SC-1 cells may be equivalent to the 12.8 kbp HindIII fragment of BALB/c, AKR, RFM/Un and SENCAR mice minus the inserted 9.0 kbp AL10 provirus. Also, on the basis of BamHI sites in the MuLV-proviral region of the AL10 DNA (Fig. 2), the detection of a 13.0 kbp BamHI fragment by the BA probe would indicate that either the AL10 provirus is not adjacent to the BA-homologous sequence in both the Swiss and C3H mice or that a loss of BamHI sites had occurred at the same or similar locations in these mice.

DISCUSSION

In the present study, all 14 recombinant DNA clones of endogenous MuLV-related sequences were found to contain the LTR gag, pol and a portion of env sequences. Furthermore, the AL10 clone examined in detail showed the typical provirus structure of cell-LTR-gag-pol-env. The presence of an EcoRI site in the env gene region is common to many endogenous non-ecotropic proviruses and has prevented us and others (44, 50) from isolating the presumed complete provirus. None of our 14 clones showed hybrdization with a specific ecotropic env probe, indicating that an endogenous ecotropic provirus (that lacks an EcoRI site) was not among these clones.

As in the other MuLV LTRs (5-9; our unpublished data for WN1802N, WN1802B, RFV and Gross Passage A viruses), the known "consensus" nucleotide sequence for eukaryotic gene expression including the "TATA" and "CAT" boxes and the polyadenylation signal are found at the expected positions in the AL10 LTR and leader regions, indicating that AL10 represents a potentially functional transcription unit in the BALB/c mouse genome. The ecotropic MuLV locus on BALB/c mouse chromosome 5, from which WN1802N virus was presumably originated, is usually not expressed in normal cells unless it is "induced" by certain genomic perturbations (51, 52). Virtually nothing is known about

the expression of AL10 and similar proviral sequences, although they constitute the majority of the MuLV-related sequences in the mouse genome. Recently, it has been demonstrated that expression of the exogenously-introduced thymidine kinase gene in the cell requires the presence of some other sequences, adjacent and upstream from the TATA and the CAT boxes, in the non-coding promoter region (53). In this regard, there are three structural features in the AL10 LTR that might be significant in relation to its promoter activity: alterations in the "enhancer" sequence area, addition of 170 bp non-homologous "insertion-like" sequence in the mid-U3 region, and duplication of the CAGCCC next to the TATA box, leading to a 6 base displacement of the CAT box further upstream. Further studies are needed to directly determine if these structural features alter the functional activity of the LTR promoter in the cell.

The most interesting feature disclosed in this study is the additional 170 bp sequence in the mid-U3 region of the AL10 LTR. The presence of short direct and inverted repeats at the termini of this sequence is similar to what has been commonly found in prokaryotic and eukaryotic transposable elements (54), although no 4-5 bp direct repeats were present immediately outside the inverted repeats as commonly found at the site of newly integrated retroviral DNA (1). Thus, if the 170 bp sequence was introduced into the AL10 LTR by an insertional or illegitimate recombination mechanism the insertion site could have been altered by evolutional changes. In independent studies, Khan and her coworkers examined by restriction endonuclease analysis various MuLV-related DNA clones from gene libraries of AKR and BALB/c strains and found that almost all of them contained an unusually large LTR of 700 bp (45); they have now determined some of the LTR nucleotide sequences and have also found insertion-like segments (Kahn, Rowe and Martin, to be published), although these sequences and the AL10 sequence remain to be compared in detail. Employing a Sau3A/Sau3A fragment (position 244 to 398) derived from this segment of the AL10 LTR as a molecular probe for gel blot analysis, we have shown that this element is also present in the non-MuLV LTR regions as interdispersed repeats in the cellular DNA of the mouse (Fig. 7), but not in other rodent (e.g. rabbit, Chinese hamster, mink and rat) or human DNAs (manuscript in preparation). No homology, however, is detected between the 170 bp sequence of the AL10 LTR and the type 2 composite of Alu-equivalent sequences in the upstream cellular region. The 170 bp sequence could thus be a new class of interdispersed short repetition sequences in the mouse. It is known from studies of transposons that an

insertion sequence element may be incorporated into the coding sequence of a gene and thereby inactivate or modify the expression of that gene (54). Further studies with the AL10 clone could help determine whether or not this type of "insertional mutation" can occur within the regulatory sequences of the cell.

Another significant observation concerns the tRNA-primer binding site of the AL10 proviral DNA. Since initiation of reverse transcription is from a tRNA primer in retroviruses, the tRNA binding site should be an essential structure if a proviral gene is to replicate through the retrovirus route. All infectious ecotropic MuLVs examined have the identical 18 base sequence complementary to the 3' 18 nucleotide sequence of proline tRNA, but the AL10 provirus has a different sequence. Thus, AL10 RNA transcripts might have a defective structure for tRNA priming function and, for this reason, might not go through the retrovirus life cycle. In this regard, it will be interesting to examine whether this type of endogenous MuLV-related provirus is capable of transposing from one chromosome locus to another within mouse cells by the direct insertional mechanism employed by transposons of prokaryotic cells. Alternatively, the primer function of the AL10 might be served by another cellular tRNA. The recently determined sequence of rat liver glutamine tRNA (47) is a close match with the AL10 primer binding site (17/18) and suggests to us that the AL10 provirus may use a glutamine isoacceptor tRNA as primer.

Inbred NFS/N, C3H, HRS/J, and California Lake Casitas feral mice and SC-1 cells apparently contain no integrated provirus at the same chromosomal site where the AL10 provirus is located in the BALB/c, AKR, RFM and SENCAR mouse genomes. It is thus possible that the establishment of this provirus integration may be as recent in evolutional age as that of the ecotropic MuLV locus.

ACKNOWLEDGMENT

We would like to express our sincere thanks to Mr. James O. Kiggans, Mr. Chong Koh, Ms. Gloria Boone and Ms. Den-Mei Yang for technical assistance, and to Dr. R. W. Tennant for valuable discussions. This research was supported by a research assistantship from the University of Tennessee (to C.Y.O.), a NIH research grant (R01-CA30308), and by the Office of Health and Environmental Research, U.S. Dept. of Energy, under contract W-7405-eng-26 with the Union Carbide Corporation.

+Present Address: National Institute of Environmental Health Sciences, P. O. Box 12233, Research Triangle Park, North Carolina 27709.

*To whom reprint requests and correspondence may be addressed at Biology Division, Oak Ridge National Laboratory, P. O. Box Y, Oak Ridge, Tennessee 37830.

REFERENCES

1. Temin, H. M. (1982) Cell 28, 3-5.
2. Varmus, H. E. (1982) Science 216, 812-820.
3. Gilboa, E., Goff, S., Shields, A., Yoshimura, F., Mitra, S. and Baltimore, D. (1979) Cell 16, 863-874.
4. Taylor, J. M., Hsu, T. W., Yeater, C. and Mason, W. S. (1979) Cold Spring Harbor Symp. Quant. Biol. 44, 1091-1096.
5. Sutcliffe, J. G., Shinnick, T. M., Verma, I. M. and Lerner, R. A. (1980) Proc. Natl. Acad. Sci. USA 77, 3302-3306.
6. Van Beveren, C., Goddard, J. G., Berns, A. and Verma, I. M. (1980) Proc. Natl. Acad. Sci. USA 77, 3307-3311.
7. Shoemaker, C., Goff, S., Gilboa, E., Paskind, M., Mitra, S. W. and Baltimore, D. (19800 Proc. Natl. Acad. Sci. USA 77, 3932-3936.
8. Dhar, R., McClement, W. L., Enquist, L. W. and Van de Woude, G. F. (1980) Proc. Natl. Acad. Sci. USA 77, 3937-3941.
9. Van Beveren, C., Rands, E., Chattopadhyay, S. K., Lowy, D. R. and Verma, I. M. (1982) J. Virol. 41, 542-556.
10. Ostrowski, M. C., Beraid, D. and Hager, G. L. (1981) Proc. Natl. Acad. Sci. USA 78, 4485-4489.
11. Fuhrman, S. A., Van Beveren, C., and Verma, I. M. (1981) Proc. Natl. Acad. Sci. USA 78, 5411-5415.
12. Blair, D. G., McClement, W. L., Oskarsson, M. K., Fischiner, P. J. and Van de Woude, G. F. (1980) Proc. Natl. Acad. Sci. USA 77, 3504-3508.
13. Blair, D. G., Oskarsson, M. K., Wood, T. G., McClements, W. L., Fischiner, P. J. and Van de Woude, G. F. (1981) Science 212, 941-943.
14. Gelb, L. D., Milstein, J. B., Martin, M. S. and Aaronson, S. A. (1973) Nature (London) New Biol. 244, 76-79.
15. Chattopadhyay, S. K., Lowy, D. R., Teich, N. M., Levine, A. S. and Rowe, W. P. (1974) Proc. Natl. Acad. Sci. USA 71, 167-171.
16. Lowy, D. R., Chattopadhyay, S. K., Teich, N. M., Rowe, W. P. and Levine, A. S. (1974) Proc. Natl. Acad. Sci. USA 71, 3555-3559.
17. Callahan, R., Benveniste, R. E., Lieber, M. M., and Todaro, G. J. (1974) J. Virol. 14, 1394-1403.
18. Steffen, D. and Weinberg, R. A. (1978) Cell 15, 1003-1010.
19. Cannani, E. and Aaronson, S. A. (1979) Proc. Natl. Acad. Sci. USA 76, 1677-1681.
20. Rowe, W. P., Hartley, J. W. and Bremner, T. (1972) Science 178, 80-862
21. Kozak, C. A. and Rowe, W. P. (1979) Science 204, 69-71.
22. Ihle, J. N., Joseph, D. R. and Domotoz, J. J. Jr. (1979) Science 204, 71-73.
23. Kozak, C. A. and Rowe, W. P. (1980) J. Exp. Med. 152, 219-228.
24. Jenkin, N. A., Copeland, N. G., Taylor, B. A. and Lee, B. K. (1981) Nature 293, 370-374.
25. Gautsch, J. W., Elder, J. H., Schindler, J., Jensen, F. C. and Lerner, R. A. (1973) Proc. Natl. Acad. Sci. USA 75, 4170-4174.
26. Benade, L. E., Ihle, J. N. and Decleve, A. (1978) Proc. Natl. Acad. Sci. USA 75, 4553-4557.

27. Boone, L. R., Myer, F. E., Yang, D. M., Kiggans, J. O., Koh, C. K., Tennant, R. W. and Yang, W. K. (1983) J. Virol. 45, 484-488.
28. Liou, R. S., Boone, L. R., Kiggans, J. O., Yang, D. M., Wang, T. W., Tennant, R. W. and Yang, W. K. (1983) J. Virol. 46, 288-292.
29. Hirt, B. (1967) J. Mol. Biol. 26, 365-369.
30. Maxam, A. M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
31. Smith, H. (1980) Methods in Enzymol. 65, 371-380.
32. Rigby, P., Dieckman, M., Rhoes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
33. Smith, H. O. and Birnstiel, M. L. (1976) Nucl. Acid Res. 3, 2387-2399.
34. Yang, W. K., Kiggans, J. O., Yang, D. M., Ou, C. Y., Tennant, R. W., Brown, A. and Bassin, R. H. (1980) Proc. Natl. Acad. Sci. USA 77, 2994-2998.
35. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
36. Alwine, J. C., Kemp, D. J. and Stark, G. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5350-5354.
37. Wahl, G. M., Stern, M. and Stark, G. R. (1979) Proc. Natl. Acad. Sci. USA 76, 3683-3687.
38. Blattner, F. R., Williams, B. G., Blechl, A., Denniston-Thompson, K., Faber, H. E., Furlong, L., Grunwald, D. J., Keifer, D. O., Moore, D. D. Schumm, J. W., Sheldon, E. L. and Smithies, O. (1977) Science 196., 161-169.
39. Benton, W. D. and Davis, R. W. (1977) Science 196, 180-182.
40. Davis, R. W., Botstein, D. and Roth, J. R. (1980) A manual for genetic engineering and advanced bacteria genetics. pp 80-83, Cold Spring Harbor Laboratory, New York.
41. Coffin, J. (1979) J. Gen. Virol. 42, 1-46.
42. Levinson, B., Khoury, G., Van de Woude, G. F. and Gruss, P. (1982) Nature 295, 568-572.
43. Rands, E., Lowy, D. R., Lander, M. R. and Chattopadhyay, S. K. (1981) Virology 108, 445-452.
44. Khan, A. S., Rowe, W. P. and Martin, M. A. (1982) J. Virol. 44, 625-636.
45. Peters, G., Harada, F., Dahlberg, J. E., Panet, A., Haseltine, W. A. and Baltimore, D. (1977) J. Virol. 21, 1031-1037.
46. Waters, L. C. and Mullin, B. C. (1977) Progr. Nucl. Acid Res. Mol. Biol. 20, 131-145.
47. Yang, J. A., Tai, L. W., Agris, P. F., Gehrke, C. W. and Wong, T.-W. (1983) Nucl. Acid Res. 11, 1991-1996.
48. Jelinek, W. r. and Schmid, C. W. (1982) Ann. Rev. Biochem. 51, 813-844.
49. Page, G. S., kSmith, S. and Goodman, H. M. (1981) Nucl. Acids Res. 9, 2087-2104.
50. Robin, R., Young, J. M., Mural, R. J., Bell, T. E., and Ihle, J. N. (1982) J. Virol. 43, 113-126.
51. Kozak, C. A. and Rowe, W. P. (1980) J. Exp. Med. 152, 219-228.
52. Ihle, J. N., Joseph, D. R. and Domotor, J. J. (1979) Science 204, 71-73.
53. McKnight, S. L. and Kinsbury, R. (1982) Science 217, 316-324.
54. Shapiro, J. A. and Cordell, B. (1982) Biol. Cell 43, 31-54.