JOURNAL OF VIROLOGY, June 1983, p. 871–880
0022-538X/83/060871-10$02.00/0
Copyright © 1983, American Society for Microbiology
Vol. 46, No. 3

# Nucleotide Sequence of the Envelope Gene of Gardner-Arnstein Feline Leukemia Virus B Reveals Unique Sequence Homologies with a Murine Mink Cell Focus-Forming Virus†

JOHN H. ELDER[1]* AND JAMES I. MULLINS[2]

*Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, California 92037,[1] and Department of Cancer Biology, Harvard School of Public Health, Boston, Massachusetts 02115[2]*

Received 21 January 1983/Accepted 4 March 1983

The nucleotide sequence of the envelope gene and the adjacent 3′ long terminal repeat (LTR) of Gardner-Arnstein feline leukemia virus of subgroup B (GA-FeLV-B) has been determined. Comparison of the derived amino acid sequence of the gp70-p15E polyprotein to those of several previously reported murine retroviruses revealed striking homologies between GA-FeLV-B gp70 and the gp70 of a Moloney virus-derived mink cell focus-forming virus. These homologies were located within the substituted (presumably xenotropic) portion of the mink cell focus-forming virus envelope gene and comprised amino acid sequences not present in three ecotropic virus gp70s. In addition, areas of insertions and deletions, in general, were the same between GA-FeLV-B and Moloney mink cell focus-forming virus, although the sizes of the insertions and deletions differed. Homologies between GA-FeLV-B and mink cell focus-forming virus gp70s is functionally significant in that they both possess expanded host ranges, a property dictated by gp70. The amino acid sequence of FeLV-B contains 12 Asn-X-Ser/Thr sequences, indicating 12 possible sites of N-linked glycosylation as compared with 7 or 8 for its murine counterparts. Comparison of the 3′ LTR of GA-FeLV-B to AKR and Moloney virus LTRs revealed extensive conservation in several regions including the "CCAAT" and Goldberg-Hogness (TATA) boxes thought to be involved in promotion of transcription and in the repeat region of the LTR. The inverted repeats that flanked the LTR of GA-FeLV-B were identical to the murine inverted repeats, but were one base longer than the latter. The region of U3 corresponding to the approximately 75-nucleotide "enhancer sequence" is present in GA-FeLV-B, but contains deletions relative to AKR and Moloney virus and is not repeated. An interesting pallindrome in the repeat region immediately 3′ to the U3 region was noted in all the LTRs, but was particularly pronounced in GA-FeLV-B. Possible roles for this structure are discussed.

The feline leukemia viruses (FeLV) are a group of horizontally transmitted retroviruses that are found associated with malignant and degenerative diseases of hematopoietic origin in domestic cats (11, 12, 18, 27, 38). As in other retrovirus systems, including murine and primate retroviruses, the mechanism(s) by which these viruses cause leukemia has not been established. Three subgroups defined by type-specific neutralization and interference assays comprise the FeLV family (40, 41). These include the following: FeLV-A, which is the most commonly isolated subgroup; FeLV-B, which is generally found in association with FeLV-A; and FeLV-C, which is rare and has been isolated along with FeLV-A and FeLV-B from cats with polycythemia. Nucleic acid hybridization has revealed extensive structural homologies between the three subgroups, showing at least 85% homology (31). More divergence is noted by T1 oligonucleotide analysis (39). FeLV subgroups vary not only in frequency of occurrence, but also in their host ranges (42) as well as in their capacity for horizontal infection when administered as discrete subtypes (28). FeLV-A is horizontally infectious, whereas FeLV-B appears to require the presence of FeLV-A for efficient horizontal spread (28). Interestingly, an apparent synergy is noted in that the incidence of viremia in multiple cat households with FeLV-AB infection is considerably greater than in households where only FeLV-A is found (27). Whether this phenomenon relates to penetration, avoidance of immune surveillance, or another mechanism is yet to be determined.

† Publication no. 2903 from the Research Institute of Scripps Clinic.

It is possible that many of the above phenomena involve the surface glycoprotein gp70, which possesses recognitive functions associated with host range and is the primary target of virus neutralization in all type C retrovirus systems tested. gp70 is a polymorphic molecule (10) and provides the determinants by which the three FeLV subtypes are defined (40, 41), as in the murine (24) and primate (23) retrovirus systems. In an attempt to relate the functional differences described above to structural changes, we have undertaken the task of nucleotide sequencing the envelope genes of the FeLV viruses. Herein, we report the nucleotide sequence and derived amino acid sequence of the Gardner-Arnstein FeLV-B (GA-FeLV-B) envelope gene which encodes gp70, as well as the 3′ long terminal repeat (LTR). Comparison of the amino acid sequence of FeLV-B gp70 with sequences of four previously published murine retrovirus gp70s has yielded valuable information regarding the localization of highly conserved regions of the molecule as well as regions that afford a structural handle to study determinants dictating viral host range.

## MATERIALS AND METHODS

**Clones.** A recombinant DNA clone of an infectious clone of GA-FeLV-B provirus in the plasmid pBR322 (designated pFGB) was utilized for these studies. The plasmid was cloned from the λ HF60 isolate, which was obtained from a Charon 4A library of *Eco*RI-cleaved DNA from GA-FeLV-B-infected human RD cells (34). The λ HF60 insert was cloned into the *Eco*RI site of pBR322. Subcloning was performed on this plasmid by utilizing an *Hin*dIII site in the 3′ portion of the *pol* gene and another outside of the integrated virus within cellular flanking sequences (Fig. 1). Ligation into the *Hin*dIII site of pBR322 yielded two plasmids, one containing a 6.5-kilobase insert comprising the 5′ portion of the virus and another (pFGB3.5) containing the envelope gene as well as the 3′ LTR of the virus (Fig. 1). The latter clone was utilized for sequence analysis.

**Isolation of recombinant DNA.** We have used a modification of the boiling procedure of Holmes and Quigley (25) to purify plasmid DNA (Sparks and Elder, manuscript in preparation). C600 cells were transformed with pFGB3.5, and colonies resistant to ampicillin and sensitive to tetracycline were selected. An overnight culture was grown from a single colony in yeast-tryptone medium in the presence of ampicillin (33 μg/ml). One-liter cultures of yeast-tryptone medium plus ampicillin were then inoculated with 5 ml of overnight culture and allowed to grow at 37°C until an optical density at 550 nm of 0.6 was attained. Chloramphenicol (0.173 g/liter) was then added, and the cultures were amplified overnight at 37°C with vigorous agitation. Cells were then harvested, washed once with M9 medium, and suspended in lysis buffer (0.05 M Tris [pH 7.8], 0.05 M EDTA, 1% Triton X-100, 8% sucrose). Lysozyme (1 mg/ml, final concentration) was added, and the solution was placed in a beaker and brought to a boil for approximately 30 s, at which time a clot formed. The solution was centrifuged at 20,000 rpm in a 60 Ti rotor for 1 h to remove chromosomal DNA and associated proteins. The supernatant was brought to 0.4 M sodium acetate and precipitated with an equal volume of isopropanol. The recovered precipitate was redissolved in 4 ml of water, the salt concentration was adjusted to 0.4 M sodium acetate, and the solution was precipitated with 2.5 volumes of ethanol. The precipitate was dissolved in 10 ml of water, and the majority of protein and associated rRNA was precipitated by the addition of 40 ml of saturated ammonium sulfate. After 1 h at 4°C, the precipitate was removed by centrifugation, and the supernatant was passed over a column (Bio-Rad Laboratories, Inc.; 1.5 by 20 cm) containing 5 g of preparative C18 silica beads (50 to 200 μm; Waters Associates) that had been pretreated with 100% methanol and equilibrated in deaerated water. The column was washed with 10 ml of water, and plasmid DNA and contaminating RNA were then eluted with 30% methanol in water. Plasmids so recovered were virtually 100% supercoiled and free of chromosomal DNA. RNA (which comprised 30 to 50% of the recovered nucleic acid) could be removed by RNase treatment, but did not interfere with sequence analysis. pFGB3.5 amplified well, and approximately 15 mg of plasmid DNA per liter was recovered by this technique. Variability in recoveries with various plasmids proved to be a function of amplification and not of experimental protocol.

**DNA sequencing.** DNA sequencing was performed essentially as described by Maxam and Gilbert (33), except that 4% formic acid was used in place of 4% pyridinium formate for the adenine plus guanine reaction.

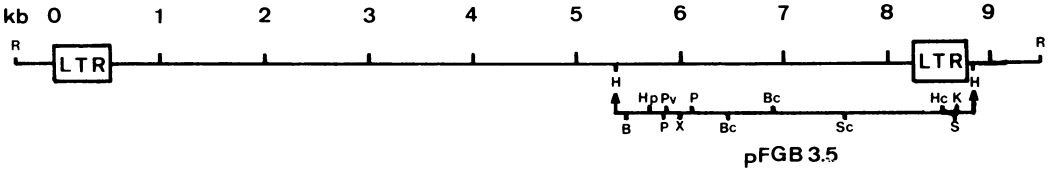


FIG. 1. Partial restriction map of pFGB3.5. The full length clone, pFGB, is contained within a 9.5-kilobase (kb) *Eco*RI (R) fragment subcloned into pBR322 from λ HF60 (34). The clone is shown with the 5′ LTR on the left. pFGB3.5 is a subclone from an *Hin*dIII (H) site at approximately 5.4 kilobases to the *Hin*dIII site outside the viral genome at 8.8 kilobases. This fragment was cloned into the *Hin*dIII site of pBR322 and utilized for subsequent sequence analyses. Abbreviations: B, *Bam*HI; Hp, *Hpa*I; P, *Pst*I; Pv, *Pvu*II; X, *Xho*I; BC, *Bcl*I; Sc, *Sac*II; Hc, *Hin*cII; S, *Sma*I; K, *Kpn*I.

## RESULTS AND DISCUSSION

**Defining the envelope gene.** The retroviral envelope gene resides at the 3′ end of the retrovirus genome and is translated from a spliced mRNA as a polyprotein containing the envelope proteins gp70 and p15E. This polyprotein is proteolytically processed into the individual proteins during virus maturation. The amino acid sequences of GA-FeLV-B gp70 and p15E have not been determined. We were, however, able to assign tentative locations for the N termini of these proteins by aligning the deduced amino acid sequence from a large open reading frame of our sequence with the amino acid sequences of Rauscher murine leukemia virus (22) and Friend murine leukemia virus (F-MuLV) (5) gp70s. In Fig. 2, the nucleotide sequence and deduced amino acid sequence of pFGB3.5 are shown from the *Pst*I site at approximately 6,100 nucleotides of the full length clone (Fig. 1) to the end of the 3′ LTR. The nucleotide sequence is shown as the sense (+) strand, and for the purposes of this report, the first nucleotide of the *Pst*I site is designated nucleotide 1. As for the murine retroviruses studied, a long open reading frame is present, which, when translated into amino acids, shows sufficient homology with previously published sequences so as to facilitate tentative location of the gene products gp70 and p15E. As with Moloney murine leukemia virus (M-MuLV) (45) and AKR ecotropic virus (Akv) (29), the coding sequence for the envelope gene is in a different reading frame from the polymerase gene, which precedes it to the 5′ side. Two initiator codons are present in GA-FeLV-B before the apparent N terminus of gp70. The first is at nucleotide 61 and would provide a 33-amino-acid leader sequence similar to those described for M-MuLV (45) and Akv (29) viruses and typical of signal sequences found on other membrane protein precursors (3). The second methionine codon is at nucleotide 158. The latter site would not provide a leader since the apparent N terminus of gp70 by alignment with the other sequences is at nucleotide 161. The nucleotide sequence in this reading frame remains open until a stop codon is reached at nucleotide 2,047. By alignment with other sequences, the N terminus of p15E is at nucleotide 1,457, which dictates a gp70 molecule 432 amino acids long and a p15E molecule comprising 197 amino acids. The precise location for processing of the polyprotein or whether amino acid residues are removed during processing is unknown for GA-FeLV-B. However, the amino acid sequence data of Chen (5) with F-MuLV gp70 indicates that a single cut is made before the N terminus of p15E, with no amino acids removed.

**Comparative analysis of gp70.** Comparison of the amino acid sequences of four murine retrovirus gp70s to FeLV-B gp70 (Fig. 3) yields considerable insight into which regions of gp70 are conserved as well as which regions are specific for a given virus or host range group. The gp70 of FeLV-B is 432 amino acids long compared with 445 for F-MuLV (5), 439 for Akv (29), 436 for M-MuLV (45), and 407 for a Moloney mink cell focus-forming virus (M-MCF) (4) recombinant virus gp70. FeLV-B gp70 contains 12 possible glycosylation sites (Asn-X-Thr/Ser; Fig. 3) as opposed to 8 for F-MuLV, 7 for Akv, 7 for M-MuLV, and 7 for M-MCF gp70s. In general, the greatest degree of conservation is in the C-terminal third of gp70, starting around position 320 of Fig. 3, although GA-FeLV-B gp70 shows more divergence in this region than do its murine counterparts, including insertion of three possible glycosylation sites and deletion of another. The N-terminal portion contains several variable regions, including three regions of insertion or deletion (the latter indicated by dashed lines) or both. These areas occur approximately every 100 positions and by definition represent areas of considerable structural tolerance to change. GA-FeLV-B and M-MCF both have large deletions around residue 100, relative to F-MuLV, Akv, and M-MuLV, which start at different residues, but terminate at the same site. Conversely F-MuLV, Akv, and M-MuLV gp70s have a 22-amino-acid deletion around residue 200 relative to GA-FeLV-B and M-MCF gp70s. The region centering around position 300 (from 270 to 320) shows the greatest degree of variability, creating "type-specific" determinants that delineate all of the gp70s shown. This region has been described as a relative "hot spot" in gp70 (5, 45) and has been suggested as being a site for recombination (45). However, the one recombinant sequenced, M-MCF (4), is not identical to the parental virus, M-MuLV, until residue 398, suggesting that at least for M-MCF, recombination occurs well past this point. The very C terminus of gp70 (positions 470 through 480) is highly variable, although many of the changes are conserved (i.e., lysine for arginine). All gp70s shown possess two basic amino acids just before p15E which would be susceptible to a trypsin-like cleavage. As stated above, results with F-MuLV (5) indicate that a single cut produces gp70 and p15E. There are at least four regions (surrounded by boxes in Fig. 3) which indicate that FeLV-B gp70 is more closely related to the endogenous murine virus gp70 (presumably xenotropic) involved in recombination to form M-MCF. These areas differ dramatically from the three ecotropic virus gp70s shown (F-MuLV, Akv, and M-MuLV), yet are highly related between FeLV-B

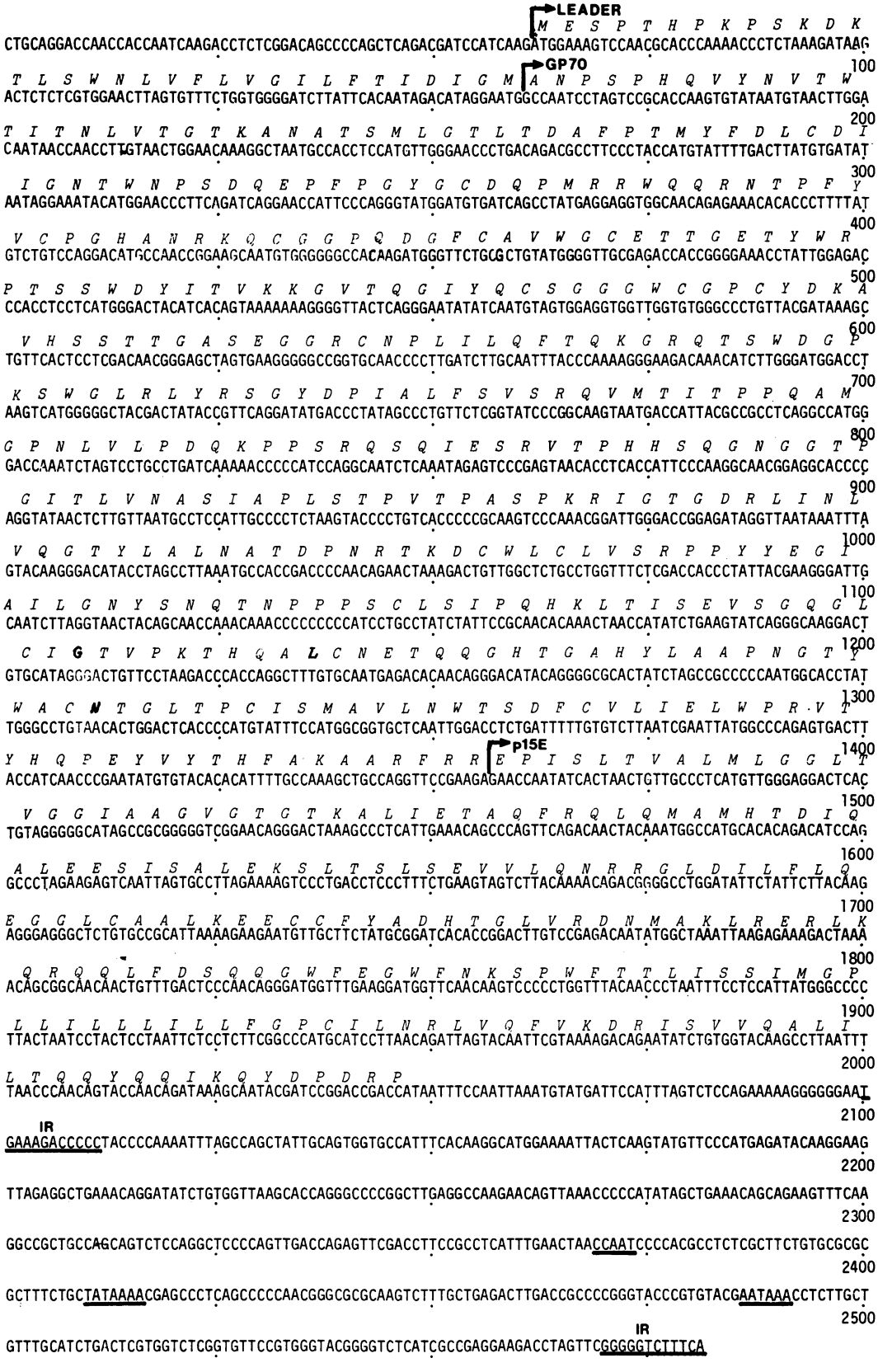


FIG. 2. Nucleotide and deduced amino acid sequence of the envelope gene and 3′ LTR of GA-FeLV-B. Nucleotide sequence is shown as the sense (+) strand (Gothic type) from a *Pst*I site at approximately 6.1 kilobases (Fig. 1) to the end of the 3′ LTR. Amino acids are shown in italics as the one-letter amino acid code. The presumed locations of leader sequence, gp70, and p15E are indicated. The "CCAAT" and Goldberg-Hogness boxes and putative sequences for polyadenylation are underlined.

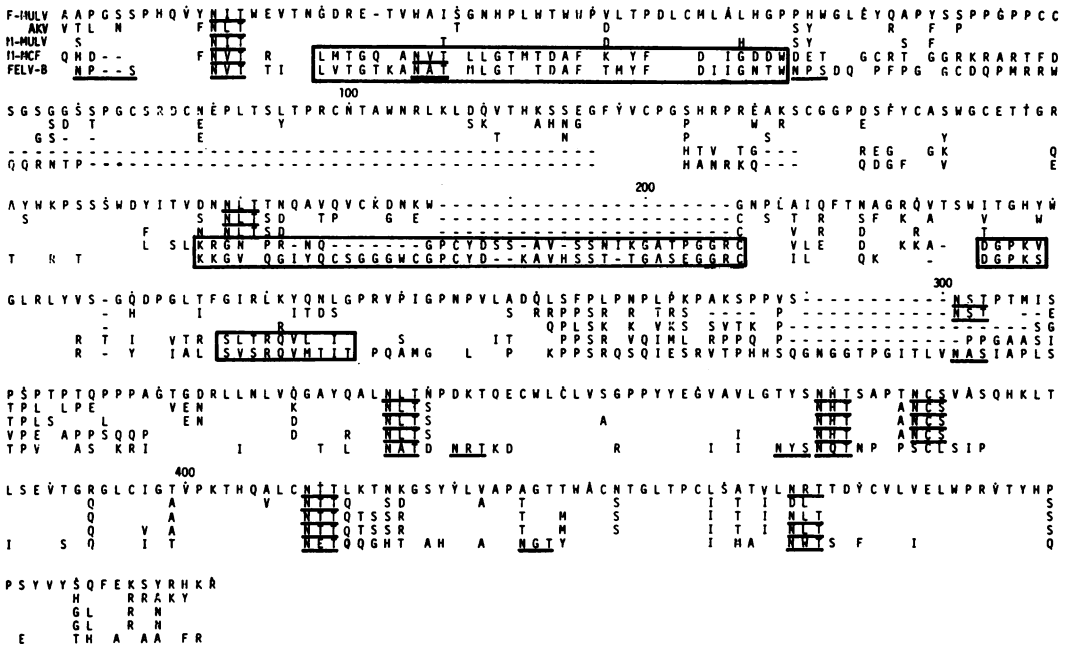

FIG. 3. Amino acid sequence comparison of FeLV-B gp70 to four murine retrovirus gp70s. Amino acids common to F-MuLV gp70 are omitted from the other sequences. Dashed lines indicate deletions, which were localized by manually aligning the nucleotide sequences of the viruses relative to highly conserved regions. Underlined amino acids show possible glycosylation sites. Boxed areas indicate regions of high homology between M-MCF and FeLV-B gp70 not shared by F-MuLV, Akv, or M-MuLV virus gp70. The data are from the following sources: F-MuLV, Chen (5); Akv, Lenz et al. (29); M-MuLV, Shinnick et al. (45); M-MCF, Bosselman et al. (4).

and M-MCF. The first region (residues 20 through 54) is near the N terminus, where FeLV-B gp70 and M-MCF show 70% (24 of 35) amino acid homology to one another, but only about 20% (8 of 35) homology to the ecotropic viruses. Within this region, both FeLV-B and M-MCF gp70s contain a glycosylation site not found in the ecotropic viruses. The second region (residues 167 through 207) shows 50% (19 of 41) homology between FeLV-B and M-MCF gp70s, with only one amino acid in common with the ecotropic virus gp70s. Here, a glycosylation site is absent in FeLV-B and M-MCF gp70s, but present in the three ecotropic virus gp70s. This area also encompasses a region of insertion for FeLV-B and M-MCF relative to the ecotropic viruses. Two other small regions (residues 225 through 229 and 247 through 256) are 80% (4 of 5) and 50% (5 of 10) homologous, respectively, between FeLV-B and M-MCF gp70, but have almost no homology with the ecotropic virus gp70s.

**Comparative analysis of p15E.** A comparative analysis of the predicted amino acid sequences of M-MuLV, Akv, and GA-FeLV-B p15E is shown in Fig. 4. The sequence of F-MuLV p15E is not available, and M-MCF p15E is inherited from the parental M-MuLV and is thus omitted. Overall, the structure of p15E is highly conserved, with 180 of 199 residues identical between M-MuLV and Akv; 154 of 199 residues identical between M-MuLV and GA-FeLV-B p15E; and 161 of 199 residues identical between Akv and GA-FeLV-B p15Es. As with gp70, p15E shows more structural divergence toward the N-terminal end of the molecule. In particular, the region 30 to 50 residues from the N terminus of p15E varies considerably in GA-FeLV-B p15E relative to M-MuLV or Akv. As for the N terminus, the C terminus of GA-FeLV-B p15E has not been determined. However, by analogy to M-MuLV p15E (S. Oroszlan, personal communication), the C terminus would be the leucine residue at position 180 of Fig. 4. As in the murine viruses, processing at this point yields a protein 180 residues long with a small piece removed from the C terminus of the polyprotein. This area, termed the R region in studies of M-MuLV (15, 47), is 17 amino acids long in GA-FeLV-B as compared to 16 and 19 amino acids long for M-MuLV and Akv, respectively. As in the murine sequences, the R region of GA-FeLV-B shows extensive divergence at the C terminus. Whether this region has functional significance is yet to be determined. However, it

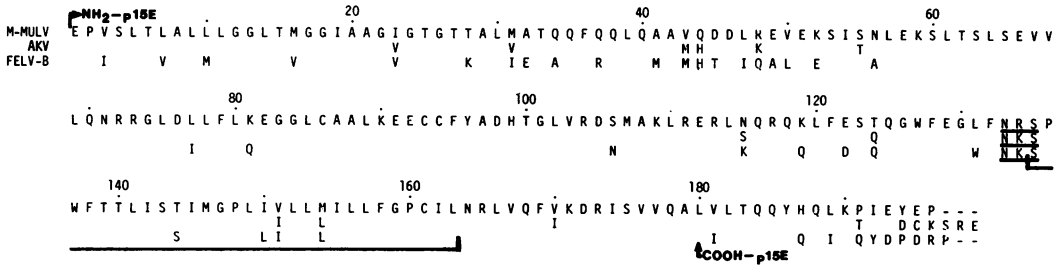


FIG. 4. Comparison of the derived amino acid sequences of GA-FeLV-B, Akv, and M-MuLV p15Es. The sequence of M-MuLV was taken from Shinnick et al. (45), and the sequence of Akv was taken from Lenz et al. (29). Only amino acids which differ from M-MuLV p15E are shown. Deletions are represented by dashed lines. A possible glycosylation site is underlined, but apparently is not utilized. The single bracket denotes a region of 29 hydrophobic amino acids that is present in all three p15Es and that possibly represents a region which traverses the viral membrane (29).

is conceivable that it would be located on the cytoplasmic side of the plasma membrane during virus maturation (29) and may represent a signal for viral core formation and subsequent budding (15, 29).

The amino acid sequences of all p15Es examined contain a single Asn-X-Ser site (underlined in Fig. 4) that could function as a site for N-linked glycosylation. However, no carbohydrate has been detected on either murine or feline p15Es. However, as pointed out by Sutcliffe et al. (45) for M-MuLV and by Lenz et al. (29) for Akv, GA-FeLV-B p15E contains an extremely hydrophobic stretch of 29 amino acids (horizontal bracket in Fig. 4) which starts with the serine residue of the putative glycosylation site. It has been proposed that these hydrophobic amino acids could form an α-helical structure which traverses the membrane (29). It is thus possible that glycosylation does not occur at this site because the third position (serine) of the recognition signal is buried in the membrane.

**Characteristics of the 3′ LTR.** Thirty-five nucleotides past the stop codon at the end of the envelope gene (Fig. 2) is a 15-nucleotide, purine-rich region that precedes the 3′ LTR at nucleotide 2,100. The GA-FeLV-B 3′ LTR is characterized by features previously described for the LTRs of other retroviruses (8, 17, 45, 48). Our sequence varies by only three nucleotides from that recently reported for the Gardner-Arnstein feline sarcoma virus LTR (17). The LTR is bordered by 12-nucleotide inverted repeats (IRs). Additionally, consensus sequences ascribed to promoter functions are present, including the five-nucleotide "CCAAT" box (16) (underlined in Fig. 2) at nucleotide 2,369 and Goldberg-Hogness box (6, 13, 36) at nucleotide 2,410. These structures are presumed to be functional for viral transcription as they exist in the identical region at the 5′ end of the integrated viral genome. The sequence "AATAAA," which is thought to be the recognition signal for polyadenylation (37), is present at nucleotide 2,486; a "CA" sequence, thought to be the acceptor for polyadenylation (37), follows at nucleotide 2,506. A comparative analysis of the GA-FeLV-B to M-MuLV (45) and Akv (29) virus LTRs is shown in Fig. 5. The sequences between the end of the envelope gene and the beginning of the LTR at the IR are included for comparison. The distance between the envelope gene and the LTR is different for each of these viruses, with GA-FeLV-B containing the most bases and Akv containing the least. Each LTR contains a highly conserved purine-rich region immediately 5′ to the IR. The minus strand through this region and including the IR can base pair to form a "cloverleaf" structure (47) and is thought to be the signal for plus strand synthesis (1, 46, 47). The 5′ IR of GA-FeLV-B is 12 nucleotides long, as compared with 11 for the murine IRs, but otherwise is identical. Overall homology (counting deletion) with GA-FeLV-B is 66 and 63% for Akv and M-MuLV, respectively, as compared with 77% between Akv and M-MuLV. Areas within the U3 regions of Akv (vertical arrows in Fig. 5; nucleotides 164 through 264) and M-MuLV (vertical arrows in Fig. 5; nucleotides 163 through 244) delineate putative "enhancer sequences" that are directly repeated in these two viruses (8, 48). A similar 72-base-pair repeat from Moloney sarcoma virus has been shown to substitute for the 72-base-pair repeat in simian virus 40 in augmenting transcription of papovavirus early genes (32). Thus, it is thought that this region is involved in enhancing transcription. The corresponding area of GA-FeLV-B contains several areas of strong homology, but also includes two deletions and is not repeated, as recently reported for feline sarcoma virus (17). It is yet to be determined whether this region of GA-FeLV possesses "enhancer" activity. The "CCAAT" (nucleotides

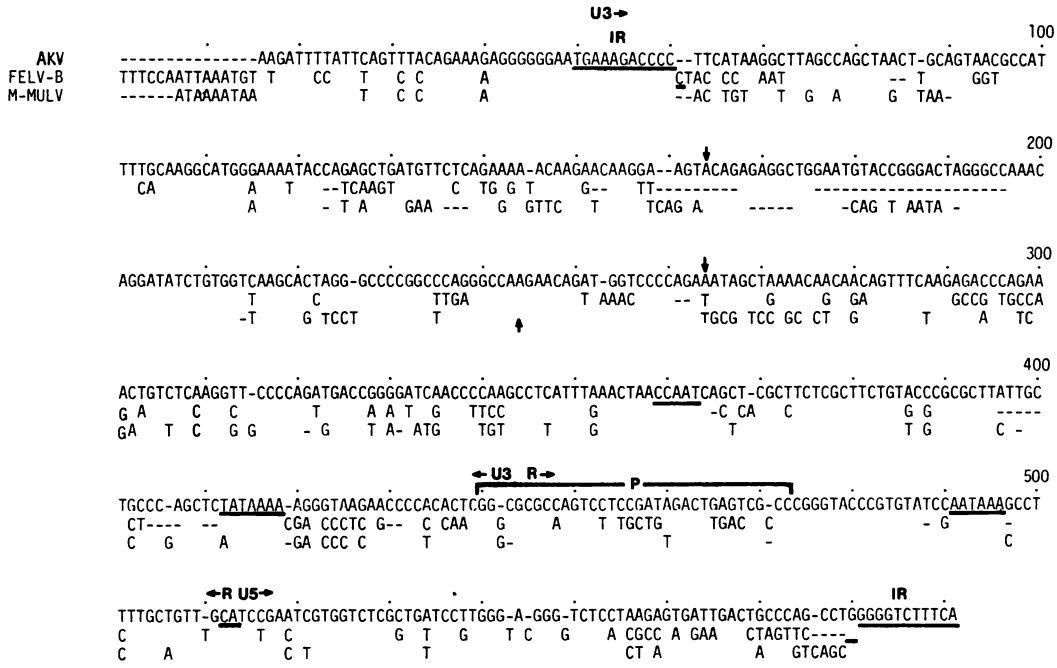


FIG. 5. Comparative analysis of the LTRs of GA-FeLV-B, Akv, and M-MuLV. The sequence of the M-MuLV LTR was taken from Shinnick et al. (45), and the sequence of the Akv LTR was taken from Van Beveren et al. (48). Common nucleotides are deleted; the deletions are indicated by dashed lines. U3 is a sequence unique to the 3' portion of the viral genome, U5 is a sequence unique to the 5' portion of the viral genome, and R is a repeat region found at both ends of the viral genome. The vertical arrows delineate regions of Akv and M-MuLV that are directly repeated in these LTRs, but not in GA-FeLV-B. P (single bracket) is the pallindromic sequence that is present in each of the LTRs.

359 through 363) and Goldberg-Hogness (nucleotides 412 through 418) boxes are conserved, although the distance between the two areas is 12 nucleotides shorter in GA-FeLV-B. Likewise, the AATAAA (nucleotides 491 through 496) thought to be the recognition signal for polyadenylation (37) is conserved as well as the CA sequence (nucleotides 512 through 513) thought to be the site of polyadenylation (37). Note that the distances from the Goldberg-Hogness box to the polyadenylation signal as well as to the presumed site of polyadenylation are conserved, although deletions occur in all three LTRs. A pallidromic sequence (P bracket in Fig. 5) was noted in all three LTRs within the R region immediately 3' to the end of U3. The structures formed by these sequences as well as the corresponding region of simian sarcoma virus (7) murine sarcoma virus (8, 48), and spleen necrosis virus (44) LTRs are shown in Fig. 6. The structures are drawn as their RNA counterparts, where presumably such structures could form and alternatives for base pairing are indicated by dashes. Note that all possess mismatch areas at the top of the structure as well as in the middle of the stem. It must be pointed out that it is not known whether such a structure could form in vivo. However, it is interesting to note that many of the nucleotides that are not conserved between these LTRs are complementary within this structure. The precise beginning of R (methyl-G "cap" site) is only known for M-MuLV. However, by analogy with M-MuLV, the border between R and U3 would fall in approximately the same location on each of these structures. Similar structures have been proposed as possible regulatory structures to facilitate termination of transcription at the 3' end utilizing further base pairing between U3 and R, which would be absent at the 5' end of the viral genome (2). In M-MuLV, six base pairs are possible between U3 and R with only a single base mismatch. Depending on the actual cap site, Akv, murine sarcoma virus, and simian sarcoma virus also have reasonable possibilities for U3 and R base pairing. However, the match in this region of FeLV-B is less profound and can only be accomodated by a three-base mismatch. FeLV-B does, however, have a substantially better base pairing capability in the long stem of the structure, with 14 of a possible 17 matches, in contrast to 10 of 17 for M-MuLV; 10 of 16 for Akv; 11 of 16 for murine sarcoma virus; and 10 of 16 for simian sarcoma virus. All of these

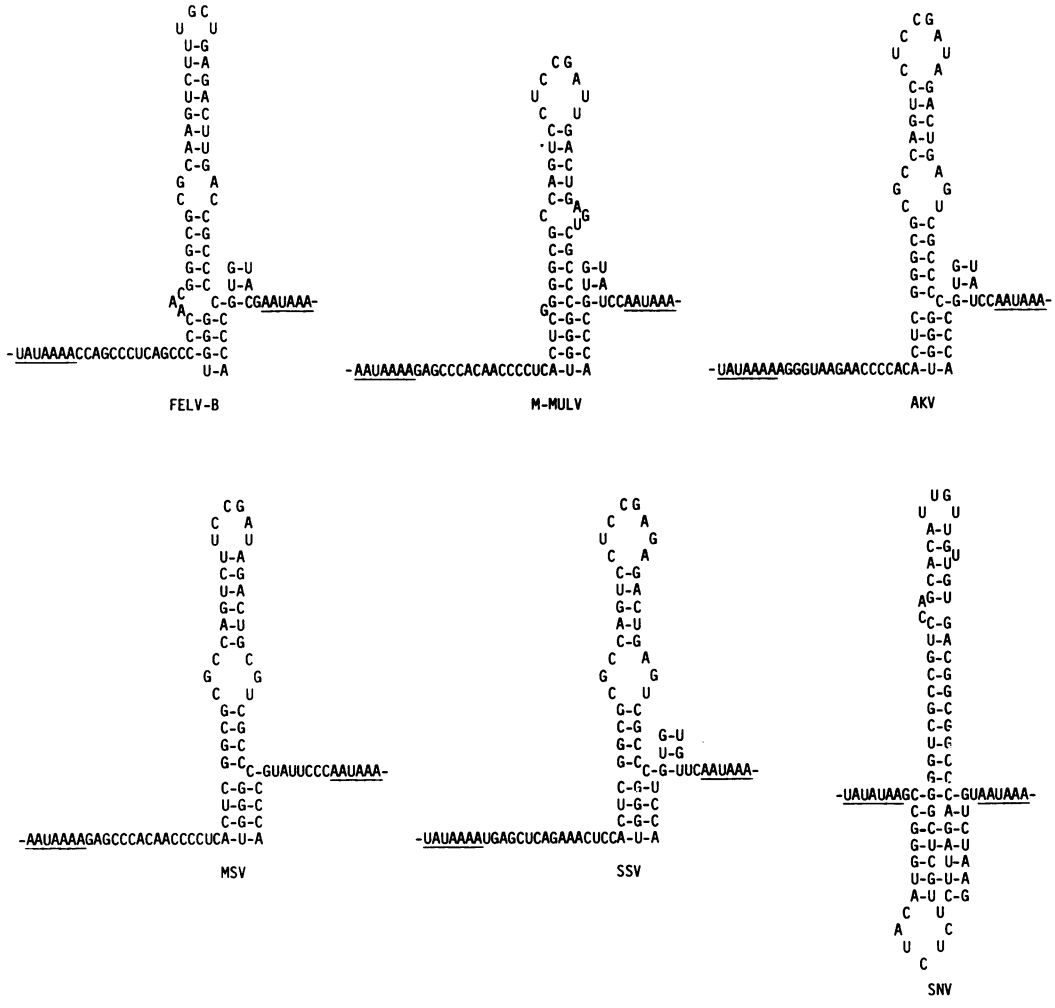


FIG. 6. Proposed structures for pallindromic sequences found in the repeat region (R) of GA-FeLV-B, M-MuLV, Akv, simian sarcoma virus (SSV), murine sarcoma virus (MSV), and spleen necrosis virus (SNV) LTRs. Structures are shown as they may exist in the viral RNA. The dashes indicate possible base pairing. The sequence of simian sarcoma virus was taken from Devare et al. (7), the spleen necrosis virus LTR sequence was taken from Shimotohno et al. (44), and the murine sarcoma virus sequence was taken from Dhar et al. (8). The putative promoter (TATA) boxes and the putative signal for polyadenylation (AAUAAA) are underlined.

structures have a relatively high G+C content (approximately 70%), particularly at the base of each stem. Although each structure has unique features, all have in common a general structure and location relative to the putative promoter (TATA) box and the signal for polyadenylation. This is also true for the avian spleen necrosis virus (Fig. 6), which can undergo much more extensive base pairing than the structures for the feline, murine, or primate retroviruses shown. As stated above, it is not known whether these structures form in vivo or whether they are functional. However, as previously noted (2), these structures are similar to those thought to be responsible for attenuation of the tryptophan operon of *Escherichia coli* (35). Furthermore, it has recently been shown that this region of the M-MuLV LTR can function in chain termination in *E. coli* (26). It is significant to note that in the latter study, cleavage of the M-MuLV LTR at the *Sma*I site still resulted in a fragment which acted as a terminator (26), indicating that base pairing between R and U3 below the main stems of the M-MuLV structures drawn in Fig. 6 is not necessary for termination. The involvement of these structures in eucaryotic transcriptional control is, however, yet to be demonstrated. Another possibility is that such structures could help facilitate the jump (20) from the 5′ to the 3′ end of the viral genome during reverse transcrip-

tion. However, Rous sarcoma virus (21, 43), which replicates by a similar jump mechanism, has no such structure within the R region; thus, formation of a secondary structure must not be mandatory.

**Functional significance of MCF homology.** The nucleotide sequence of GA-FeLV-B envelope gene and 3′ LTR indicate that the overall structures of the FeLV and MuLV envelope genes are quite similar. The data indicate an ancestral relationship compatible with function similarities (see below) between FeLV-B envelope and the envelope gene of the endogenous murine virus involved in MCF-type (19) recombinations. The present comparative analysis of gp70 sequences provides a framework for further studies of how gp70 binds to cells to facilitate penetration and subsequent viral expression. Clearly, there are several regions that can tolerate considerable change (large insertions or deletions) without abrogating the function or overall structural integrity of the virus particle. Furthermore, the changes in structure correlate with the host range characteristics dictated by each respective gp70. Both the GA-FeLV-B and M-MCF viruses can infect cells of various species, whereas the ecotropic viruses can only infect mouse cells. The regions of homology between M-MCF and FeLV-B are within the area of the MCF virus that is inherited via recombination from an endogenous murine virus that is highly related to certain xenotropic viruses (19). The MCF viruses, however, retain their ability to infect mouse cells as well as cells of other species (19), a trait not shared by GA-FeLV-B (42). At this point, it is not clear that all FeLV host range restrictions relate to cellular binding; they may involve restriction at another level. From a standpoint of gp70 structure, it is difficult to assign a single linear stretch of amino acids which conveniently satisfies all criteria for cellular recognition. It becomes an issue of how much change is sufficient to alter the binding site. Possibly one amino acid change is all that is required. We are pursuing this issue by chemically synthesizing peptides dictated by the deduced amino acid sequence of GA-FeLV-B and M-MCF gp70. By making antibodies to these peptides, we produce a panel of specific reagents which allow us to determine what areas of the molecule are involved in neutralization. Although neutralization may not constitute direct binding to the actual binding site, we should be able to determine if binding to different species cells is controlled by unique regions of the molecule. The synthetic peptide approach has been useful in studies of other viruses including influenza (14) and hepatitis (30) virus.

In addition, the differences in glycosylation between FeLV-B gp70 and the other viral gp70s is of interest. The nucleotide sequence dictates 12 possible glycosylation sites for FeLV-B gp70 versus 7 or 8 for the murine gp70s. Cleavage with a unique endoglycosidase, endo F (9), which cleaves both high mannose and complex N-linked carbohydrate chains, is suggestive that all the predicted glycosylation sites of FeLV-B are utilized (data not shown). However, due to insufficient resolution in our gel system, we were unable to count the number of carbohydrate chains with confidence and are currently pursuing higher resolution. How carbohydrate chains affect the overall structure of the molecule as well as immunological responsiveness is a point of continued research.

## ACKNOWLEDGMENTS

We thank Douglas Bingham, Paula Eisenhart, and Shirley Johnson for excellent technical assistance. We also thank Norman Davidson for support of J.I.M. during the initial characterization of clones utilized in this study. Thomas Shinnick, J. Gregor Sutcliffe, Michael Wilson, James Gautsch, and Richard Lerner are gratefully acknowledged for valuable discussions and technical advice.

J.I.M. was a postdoctoral fellow of the National Institutes of Health during a portion of this work and is presently a Scholar of the American Cancer Society, Massachusetts Division. This research was supported in part by Public Health Service grants CA25803 and CA25533 to J.H.E. from the National Institutes of Health and by a grant from the American Cancer Society to J.I.M.

## LITERATURE CITED

1. **Baltimore, D., E. Gilboa, E. Rothenberg, and F. Yoshimura.** 1978. Production of a discrete, infectious, double-stranded DNA by reverse transcription in virions of Moloney murine leukemia virus. Cold Spring Harbor Symp. Quant. Biol. **43**:869–874.
2. **Benz, E. W., R. M. Wydro, B. Nadal-Ginard, and D. Dina.** 1980. Moloney murine sarcoma proviral DNA is a transcriptional unit. Nature (London) **288**:665–669.
3. **Blobel, G., and B. J. Dobberstein.** 1975. Transfer of proteins across membranes. J. Cell Biol. **67**:835–851.
4. **Bosselman, R. A., F. Van Straaten, C. Van Beveren, I. M. Verma, and M. Vogt.** 1982. Analysis of the *env* gene of a molecularly cloned and biologically active Moloney mink cell focus-forming (MCF) proviral DNA. J. Virol. **44**:19–31.
5. **Chen, R.** 1982. Complete amino acid sequence and glycosylation sites of glycoprotein gp71A of Friend murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **79**:5788–5792.
6. **Corden, J., B. Wasylyk, A. Buchwalder, P. Sassone-Corsi, C. Kedinger, and P. Chambon.** 1980. Promoter sequences of eukaryotic protein-coding genes. Science **209**:1406–1414.
7. **Devare, S., E. P. Reddy, J. D. Law, and S. A. Aaronson.** 1982. Nucleotide sequence analysis of the long terminal repeat of integrated simian sarcoma virus: evolutionary relationship with other mammalian retroviral long terminal repeats. J. Virol. **42**:1108–1113.
8. **Dhar, R., W. L. Clements, L. W. Enquist, and G. F. Van De 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.
9. **Elder, J. H., and S. A. Alexander.** 1982. Endo-β-N-acetylglucosominidase F: endoglycosidase from *flavobacterium meningosepticum* which cleaves both "high mannose" and "complex" type glycoproteins. Proc. Natl. Acad. Sci. U.S.A. **29**:4540–4544.

10. **Elder, J. H., F. C. Jensen, M. L. Bryant, and R. A. Lerner.** 1977. Polymorphism of the major envelope glycoprotein (gp70) of murine C-type viruses: virion associated and differentiation antigens encoded by a multi-gene family. Nature (London) **267**:23–28.
11. **Essex, M.** 1975. Horizontally and vertically transmitted oncornaviruses of cats. Adv. Cancer Res. **21**:175–248.
12. **Essex, M., S. M. Cotter, A. H. Sliski, W. D. Hardy, J. R. Stephenson, S. A. Aaronson, and O. Jarrett.** 1977. Horizontal transmission of feline leukemia cluster household. Int. J. Cancer **19**:90–96.
13. **Estratiadis, A., J. W. Posakony, T. Maniatis, R. M. Lawn, C. O'Connell, R. A. Spritz, J. K. DeRiel, B. G. Forget, S. M. Weissman, J. L. Slightom, A. E. Blechl, O. Smithies, F. E. Baralle, C. C. Shoulders, and N. J. Proudfoot.** 1980. The structure and evolution of the human β-globin gene family. Cell **21**:653–668.
14. **Green, N., H. Alexander, A. Olson, S. Alexander, T. M. Shinnick, J. G. Sutcliffe, and R. A. Lerner.** 1982. Immunogenic structure of the influenza virus hemagglutinin. Cell **28**:477–487.
15. **Green, N., T. Shinnick, O. Witte, A. Ponticelli, J. G. Sutcliffe, and R. A. Lerner.** 1981. Sequence-specific antibodies show that maturation of Moloney leukemia virus envelope polyprotein involves removal of a COOH-terminal peptide. Proc. Natl. Acad. Sci. U.S.A. **78**:6023–6027.
16. **Grosschedl, H., and M. L. Birnstiel.** 1980. Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. Proc. Natl. Acad. Sci. U.S.A. **77**:1432–1436.
17. **Hampe, A., M. Gobet, J. Even, C. J. Sherr, and F. Galibert.** 1983. Nucleotide sequences of feline sarcoma virus long terminal repeats and 5′ leaders show extensive homology to those of other mammalian retroviruses. J. Virol. **45**:466–472.
18. **Hardy, W. D., L. J. Old, P. W. Hess, M. Essex, and S. Cotter.** 1973. Horizontal transmission of feline leukemia virus. Nature (London) **144**:266–269.
19. **Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe.** 1977. A new class of murine leukemia virus associated with development of spontaneous lymphomas. Proc. Natl. Acad. Sci. U.S.A. **74**:789–792.
20. **Haseltine, W. A., J. M. Coffin, and T. C. Hageman.** 1979. Structure of products of the Moloney murine leukemia virus endogenous DNA polymerase reaction. J. Virol. **30**:375–383.
21. **Haseltine, W. A., A. M. Maxam, and W. Gilbert.** 1977. Rous sarcoma virus genome is terminally redundant: the 5′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:989–993.
22. **Henderson, L. E., T. D. Copeland, G. W. Smythers, H. Marquardt, and S. Oroszlan.** 1978. Amino terminal amino acid sequence and carboxyl-terminal analysis of Rauscher murine leukemia virus glycoproteins. Virology **85**:319–322.
23. **Hino, S., J. R. Stephenson, and S. A. Aaronson.** 1975. Antigenic determinants of the 70,000 molecular weight glycoprotein of woolly monkey type-C RNA virus. J. Immunol. **115**:922–927.
24. **Hino, S., J. R. Stephenson, and S. A. Aaronson.** 1976. Radioimmunoassays for the 70,000-molecular weight glycoproteins of endogenous mouse type C viruses: viral antigen expression in normal mouse tissues and sera. J. Virol. **18**:933–941.
25. **Holmes, D. S., and M. Quigley.** 1981. A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**:193–197.
26. **Honigman, A., and A. Panet.** 1983. A unique sequence in murine leukemia virus long terminal repeat functions as a termination signal for transcription in *Escherichia Coli*. J. Virol. **45**:456–461.
27. **Jarrett, O., W. D. Hardy, M. C. Golden, and D. Hay.** 1978. The frequency of occurence of feline leukemia virus subgroups in cats. Int. J. Cancer **21**:334–337.
28. **Jarrett, O., and P. H. Russell.** 1978. Differential growth and transmission in cats of feline leukemia viruses of subgroups A and B. Int. J. Cancer **21**:466–472.
29. **Lenz, J., R. Crowther, A. Straceski, and W. A. Haseltine.** 1982. Nucleotide sequence of the AKV *env* gene. J. Virol. **42**:519–529.
30. **Lerner, R. A., N. Green, H. Alexander, F.-T. Liu, J. G. Sutcliffe, and T. M. Shinnick.** 1981. Chemically synthesized peptides predicted from the nucleotide sequence of the hepatitis B virus genome elicit antibodies reactive with the native envelope protein of Dane particles. Proc. Natl. Acad. Sci. U.S.A. **78**:3403–3407.
31. **Levin, R., S. K. Ruscetti, W. P. Parks, and E. M. Scolnick.** 1976. Expression of feline type-C virus in normal and tumor tissues of the domestic cat. Int. J. Cancer **18**:661–671.
32. **Levinson, B., G. Khoury, G. Van De Woude, and P. Gruss.** 1982. Activation of SV40 genome by 72-base-pair tandem repeats of Moloney sarcoma virus. Nature (London) **295**:568–572.
33. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labelled DNA with base-specific chemical cleavage. Methods Enzymol. **65**:499–560.
34. **Mullins, J. I., J. W. Casey, M. O. Nicolson, K. B. Burck, and N. Davidson.** 1981. Sequence arrangement and biological activity of cloned feline leukemia virus proviruses from a virus-productive human cell line. J. Virol. **38**:688–703.
35. **Oxender, D. L., G. Zurowski, and C. Yonofsky.** 1979. Attenuation in the *Escherichia Coli* tryptophan operon: role of RNA secondary structure involving the tryptophan codon region. Proc. Natl. Acad. Sci. U.S.A. **76**:5524–5528.
36. **Proudfoot, N. J.** 1979. Eukaryotic promoters? Nature (London) **279**:376.
37. **Proudfoot, N. J., and G. G. Brownlee.** 1974. Sequence of the 3′ end of the globin mRNA shows homology with immunoglobulin light chain mRNA. Nature (London) **252**:359–362.
38. **Rickard, C. G., J. E. Post, F. Noronka, and L. M. Barr.** 1969. A transmissible virus-induced lymphocytic leukemia of the cat. J. Natl. Cancer Inst. **42**:987–1013.
39. **Rosenberg, Z. F., F. S. Pedersen, and W. A. Haseltine.** 1980. Comparative analysis of the genomes of feline leukemia viruses. J. Virol. **35**:542–546.
40. **Sarma, P., and T. Log.** 1973. Subgroup classification of feline leukemia and sarcoma viruses by viral interference and neutralization tests. Virology **54**:160–169.
41. **Sarma, P. S., and T. Log.** 1971. Viral interference in feline leukemia-sarcoma complex. Virology **44**:352–358.
42. **Sarma, P. S., T. Log, D. Jain, P. R. Hill, and R. J. Huebner.** 1975. Differential host range of viruses of feline leukemia-sarcoma complex. Virology **64**:438–446.
43. **Schwartz, D. E., P. C. Zamecnik, and H. L. Weith.** 1977. Rous sarcoma virus genome is terminally redundant: the 3′ sequence. Proc. Natl. Acad. Sci. U.S.A. **74**:994–998.
44. **Shimotohno, K., S. Mizutani, and H. M. Temin.** 1980. Sequence of retrovirus provirus resembles that of bacterial transposable elements. Nature (London) **285**:550–554.
45. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
46. **Sorge, J., and S. H. Hughes.** 1982. Polypurine tract adjacent to the U3 region of the Rous sarcoma virus genome provides a *cis*-acting function. J. Virol. **43**:482–488.
47. **Sutcliffe, J. G., T. M. Shinnick, I. M. Verma, and R. A. Lerner.** 1980. Nucleotide sequence of Moloney leukemia virus: 3′ end reveals details of replication, analogy to bacterial transposons and an unexpected gene. Proc. Natl. Acad. Sci. U.S.A. **77**:3302–3306.
48. **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 pre-integration site. J. Virol. **41**:542–556.