

# Nucleotide Sequence of AKV Murine Leukemia Virus

WINSHIP HERR†

*Laboratory of Molecular Biology, Medical Research Council, Cambridge CB2 2QH, England*

Received 22 July 1983/Accepted 4 October 1983

AKV is an endogenous, ecotropic murine leukemia virus that serves as one of the parents of the recombinant, oncogenic mink cell focus-forming viruses that arise in preleukemic AKR mice. I report the 8,374-nucleotide-long sequence of AKV, as determined from the infectious molecular clone AKR-623. The 5'-leader sequence of AKV extends to nucleotide 639, after which lies a long open reading frame encoding the *gag* and *pol* gene products. The reading frame is interrupted by a single amber codon separating the *gag* and *pol* genes. The *pol* gene overlaps the *env* gene within the 3' region of the AKV genome. The nucleotide sequence of the 5' region of AKV reveals the following features. (i) The 5'-leader sequence lacks any AUG codon to initiate translation of gPr80<sup>gag</sup>, suggesting that gPr80<sup>gag</sup> is not required for the replication of AKV. (ii) A short portion of the leader region diverges in sequence from the closely related Moloney murine leukemia virus and appears to be related to a sequence highly repeated in eucaryotic genomes. (iii) As in Moloney murine leukemia virus, there is a potential RNA secondary structure flanking the amber codon that separates the *gag* and *pol* genes. This structure might function as a regulatory protein binding site that controls the relative levels of synthesis of the *gag* and *pol* precursors. The nucleotide sequence of the 3' region of AKV is compared with sequences reported previously from both infectious and noninfectious molecular clones of AKV.

The murine leukemia virus (MLV) AKV resides in the germ line of AKR mice and is implicated in the high incidence of thymic leukemias that arise when these mice are 6 to 9 months old. AKV is expressed in AKR mice from birth (41) and is distinguished from the other MLV endogenous to AKR mice because it is ecotropic; i.e., it can infect mouse cells. Low-leukemic strains of mice (e.g., NIH/Swiss) which have acquired a germ line AKV provirus from AKR mice become viremic and subsequently contract leukemia, albeit with a considerably longer latency period (39, 40). Nevertheless, AKV itself does not appear to be directly oncogenic because it is unable to induce leukemia in healthy mice (18, 29).

During the development of AKR mice, AKV recombines with germ line noncotropic MLV to produce recombinant MLVs (4, 5, 9, 17, 19, 23, 38) not found in the germ line of AKR mice (3, 17, 36). The recombinants, called MCF, appear in the preleukemic thymus (14) and are thought to be the proximal oncogenic agent in AKR leukemogenesis because they can accelerate the onset of leukemia if injected into young AKR mice (6, 31), and they are found integrated in the genomes of leukemic thymocytes (3, 16, 36). Because the noncotropic parent of MCF viruses has not been identified, the noncotropic sequences within MCF genomes have been localized by comparison to AKV.

The nucleotide sequence of the 3' region of the AKV genome has been reported previously. Lenz et al. (21) and Van Beveren et al. (56) determined the nucleotide sequence of the envelope (*env*) gene and the long terminal repeat (LTR), respectively, from an infectious molecular clone of AKV called AKR-623 (22). We previously reported the nucleotide sequence of the majority of the *pol* gene and the entire *env* gene from a noninfectious molecular clone of AKV (15). These reports concentrated on the 3' region of AKV because the *env* gene and LTR sequences of oncogenic

MCF viruses consist of a complicated mosaic of ecotropic and noncotropic MLV sequences (4, 19, 23, 38). Nevertheless, sequences within the 5' half of the MCF genome frequently appear to be derived from the noncotropic parent MLV (4, 23). To define fully the AKV genomic structure, I have determined the nucleotide sequence of the 5' region of AKV and redetermined the nucleotide sequence of the 3' region from the infectious AKR-623 clone.

## MATERIALS AND METHODS

**Sequence determination.** The nucleotide sequence was derived from the infectious molecular clone of AKV, AKR-623 (22), generously provided by D. Lowy and S. Chattopadhyay. The nucleotide sequence was determined by the chain terminator method of Sanger et al. (43), after cloning random fragments into the single-stranded bacteriophage M13 (1, 42). The sequence of the 5' region of AKV was derived from the large 4.5-kilobase *Xma*I (*Sma*I) fragment spanning this region. This fragment was purified by agarose gel electrophoresis, concatenated and circularized with T4 ligase, and sheared randomly by sonication (7). The sonicated ends were repaired with T4 polymerase and nucleotide triphosphates. Fragments 400 to 1,000 nucleotides long were selected by agarose gel electrophoresis and cloned into the *Sma*I site of the bacteriophage M13 vector mp8 (24). A total of 182 M13 clones were randomly chosen for sequencing. The inserts were sequenced by the chain terminator method, and the samples were analyzed by electrophoresis through buffer gradient gels as described by Biggin et al. (2). The sequences were compiled and ordered by computer, using the automatic DB system (51). Each nucleotide was sequenced an average of six times, and the entire region was sequenced at least once on both DNA strands.

The nucleotide sequence of the 3' region of AKV was redetermined by using the same strategy as above except that the AKR-623 *Xho*I/*Eco*RI fragment stretching from the middle of the AKV genome into the 3' cellular DNA sequences was used to generate the random M13 clones; in general, only one of the DNA strands was sequenced to

† Present address: Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724.



ArgValGluAspIleHisProThrValProAsnProTyrAsnLeuLeuSerGlyLeuProProSerHisArgTrpTyrThrValLeuAsp pol  
 ACGGGGTGAGACATCCACCCACCGTGCCAATCCTTATAACCTCTAAGTGGGCTCCACCGCTCCACCGGTGGTATACTGCTGCTG  
 2980 2990 -----Oligo. 36----- 3030 3040 3050 3060

LeuLysAspAlaPhePheCysLeuArgLeuHisProThrSerGlnProLeuPheAlaPheGluTrpArgAspProGlyMetGlyIleSer pol  
 ATTTAAGGATGCTTTTTTCGCTGAGACTCCACCCACAGTCAGCTCTCTTCGCTTCGAGTGGAGATCCAGGATGGGAATCT  
 3070 3080 3090 3100 3110 3120 3130 3140 3150

GlyAlaLeuThrTrpThrArgLeuProGlnGlyPheLysAsnSerProThrLeuPheAspGluAlaLeuHisArgAspLeuAlaAspPhe pol  
 CAGGACAATTAACCTGGACAGACTCCACAGGGTTTCAAAAACAGTCCACCTGTTTGTATGAGGACTACACAGAGACTAGCAGACT  
 3160 3170 3180 3190 3200 3210 3220 3230 3240

ArgIleGlnHisProAspLeuIleLeuLeuGlnTyrValAspAspIleLeuLeuAlaAlaThrSerGluLeuAspCysGlnGlnGlyThr pol  
 TCGGATCCAGCACAGACTTGTCTGCTCAGTACGTGGATGACATACTACTGGCCGCACTTCGAGCTCGACTGCCAACAGTA  
 3250 3260 3270 3280 3290 3300 3310 3320 3330

ArgAlaLeuLeuThrLeuGlyAsnLeuGlyTyrArgAlaSerAlaLysLysAlaGlnLeuCysGlnLysGlnValLysTyrLeuGly pol  
 CTGGCCCTGTACTAACCTAGGAACCTCGGCTTCGGCTCGCCCAAGAAAGCCCACTTTCGAGAACAGCTCAAGTATCTGG  
 3340 3350 3360 3370 3380 3390 3400 3410 3420

TyrLeuLeuLysGluGlyGlnArgTrpLeuThrGluAlaArgLysGluThrValMetGlyGlnProThrProLysThrProArgGlnLeu pol  
 GGTATCTCTAAAAGCGCTCAGAGATGGCTGACTGAGCCAGAAAGAGACTGTATGGGCGACTACTCCGAGACCCCTCGACATC  
 3430 3440 3450 3460 3470 3480 3490 3500 3510

ArgGluPheLeuGlyThrAlaGlyPheCysArgLeuTrpIleProGlyPheAlaGluMetAlaAlaProLeuTyrProLeuThrLysThr pol  
 TAAGGGAGTTTCTAGGACGGCAGGCTTCTGCTGCTGATCCCTGGTTTGGGAAATGGACCCCTTCTATCCTTACCAAAA  
 3520 3530 3540 3550 3560 3570 3580 3590 3600

GlyThrLeuPheAsnTrpGlyProAspGlnGlnLysAlaTyrGlnGluIleLysGlnAlaLeuLeuThrAlaProAlaLeuGlyLeuPro pol  
 CCGGACTCTTCAATTGGGGCCAGCAGCAAAAGCCCTATCAAGAAATCAAAACAGGCTCTCAACTGCCCGCCCTGGGATTC  
 3610 3620 3630 3640 3650 3660 3670 3680 3690

AspLeuThrLysProPheGluLeuPheValAspGluLysGlnGlyTyrAlaLysGlyValLeuThrGlnLysLeuGlyProTrpArgArg pol  
 CAGATTGACTAAGCCCTTGAACCTTTTGTGACGAGAACAGGCTACGCCAAAGGCTCTCAACGCAAAAGCTGGGACTGGCCCT  
 3700 3710 3720 3730 3740 3750 3760 3770 3780

ProValAlaTyrLeuSerLysLysLeuAspProValAlaAlaGlyTrpProProCysLeuArgMetValAlaIleAlaValLeuThr pol  
 GGCCGCTGGCTACTCTCCAAAAGCTAGATCCAGTGGCAGCTGGTGGCCCTTGTCTACGGATGGTAGCAGCCATTCGCTTGA  
 3790 3800 3810 3820 3830 3840 3850 3860 3870

LysAspAlaGlyLysLeuThrMetGlyGlnProLeuValIleLeuAlaProHisAlaValGluAlaLeuValLysGlnProProAspArg pol  
 CAAAAGATGGAGCAAGCTAATATGGGACAGCGCTAGTCACTCTGGCCCGCATCCAGTAGAGGACTGGTCAAGCAAGCCCTGACC  
 3880 3890 3900 3910 3920 3930 3940 3950 3960

TrpLeuSerAsnAlaArgMetThrHisTyrGlnAlaMetLeuLeuAspThrAspArgValGlnPheGlyProValAlaLeuAsnPro pol  
 GCTGATCTCAACCCCGCATGACCCACTACAGCAATGCTCTAGACCTGACCGATTTCAGTTCCGACCGTGGTGGCCCTCAATC  
 3970 3980 3990 4000 4010 4020 4030 4040 4050

AlaThrLeuLeuProGluGlyAlaProHisAspCysLeuGluIleLeuAlaGluThrHisGlyThrArgProAspLeuThr pol  
 CTGCCACTTACTCCTCCCGGAAAGAGGAGCCCACTGATGGCTCGAGATCTGGCTGAAACCGATGGAAACAGCCGATCTCA  
 -----Oligo. 7----- 4080 4090 4100 4110 4120 4130 4140

AspGlnProIleProAspAlaAspHisThrTrpTyrThrAspGlySerSerPheLeuGlnGluGlyGlnArgLysAlaGlyAlaAlaVal pol  
 CCGACCCCACTCCAGCCCGCAGCAGCAGTATCCGATGGGAGCAGCTTTTTCGAAAGGACAGCGAAAGCTGGGGCAGCGG  
 4150 4160 4170 4180 4190 4200 4210 4220 4230

ThrThrGluThrGluValIleTrpAlaArgAlaLeuProAlaGlyThrSerAlaGlnArgAlaGluLeuIleAlaLeuThrGlnAlaLeu pol  
 TGACCACTGAGACCGAGGTAATCTGGGCAAGGGCACTCCGGCTGGAACTCTGCCCAAGCGGCGAAGCTAGCAGTCACTCAAGCT  
 4240 4250 4260 4270 4280 4290 4300 4310 4320

LysMetAlaGluGlyLysArgLeuAsnValTyrThrAspSerArgTyrAlaPheAlaThrAlaHisIleHisGlyGluIleTyrArgArg pol  
 TGAAGTGGCAGAAAGTAAAGGCTAAACCTTTCACTGACAGCCGATGCTTTCGACCGCCATATCCATGGAGAGACTATAGGA  
 4330 4340 4350 4360 4370 4380 4390 4400 4410

ArgGlyLeuLeuThrSerGluGlyArgGluIleLysAsnLysSerGluIleLeuAlaLeuLysAlaLeuPheLeuProLysArgLeu pol  
 GCGGAGGTTGCTAACCTCAGAGGTTAGAAATCAAAAAGAGCCGATCTCGCTTACTGAAAGCTCTTTTTCGCAAGAGAC  
 4420 4430 4440 4450 4460 4470 4480 4490 4500

SerIleIleHisCysLeuGlyHisGlnLysGlyAspSerAlaGluAlaArgLysAsnArgLeuAlaAspGlnAlaAlaArgGluAlaAla pol  
 TCAGTATATCCACTGCTGGGCATCAAAAAGGAGATAGTCCGAGGCTAGGGGCAAGCCCTAGCAGACCAAGCCCGGAGGAC  
 4510 4520 4530 4540 4550 4560 4570 4580 4590

IleLysThrProProAspThrSerThrLeuLeuIleGluAspSerThrProTyrThrProAlaTyrPheHisTyrThrGluThrAspLeu pol  
 CCATAAAGACGCTCCAGATACACTCCTCTTATAGAGATTCAACCCCATATCCGCTGCTATTTCATTAATCAAGACAGATC  
 4600 -----Oligo. 6----- 4640 4650 -----Oligo. 24----- 4680

LysLysLeuArgGluLeuGlyAlaThrTyrAsnGlnSerLysGlyTyrTrpValPheGlnGlyLysProValMetProAspGlnPheVal pol  
 TAAAGAACTCAGAGACTTGGGCCCACTATAACAGCAAGGATAGTGGCTTCCAAGGCAAGCCGCTGATGCCGCAATTTG  
 4690 4700 4710 4720 4730 4740 4750 4760 4770

PheGluLeuLeuAspSerLeuHisArgLeuThrHisLeuGlyTyrGlnLysMetLysAlaLeuLeuAspArgGlyGluSerProTyrTyr pol  
 TATTGAATGTTAGACTCACTCCACCGCTCCACCTCGGCTACCAGAAATGAAGCACTCTTGACAGAGCAGCAAAAGCCCTACT  
 4780 4790 4800 4810 4820 4830 4840 4850 4860

MetLeuAsnArgAspLysThrLeuGlnTyrValAlaAspSerCysThrValCysAlaGlnValAsnAlaSerLysAlaLysIleGlyAla pol  
 ACATGCTAAACCGGCAAAAACCTCCAATGTGGCAGATCTCTGACGGCTCTGTCGCAAGTAAATCCAGCAAGCTAAATCGGGC  
 4870 -----Oligo. 4----- 4900 4910 4920 4930 4940 4950

GlyValArgValArgGlyHisArgProGlySerHisTrpGluIleAspPheThrGluValLysProGlyLeuTyrGlyTyrLysTyrLeu pol  
 CAGGAGTACAGTACAGGACATCGGCTCCAGCTCCATTTGGAGATCCATTTTACAGAGTCAAGCCGCTGATGGGTACAGACTACC  
 4960 4970 4980 4990 5000 5010 5020 5030 5040

LeuValPheValAspThrPheSerGlyTrpValGluAlaPheProThrLysArgGluThrAlaArgValValSerLysLysLeuLeuGlu pol  
 TCGTGTATTCTGTGACACTCTCTGCTTGGGTGGAAGCTTTCAACCAAGAGAGAAAAGCAGCAGACTCCTGTCGAAATGCTGGC  
 5050 5060 5070 5080 5090 5100 5110 5120 5130

GluIlePheProArgPheGlyMetProGlnValLeuGlySerAspAsnGlyProAlaPheThrSerGlnValSerGlnSerValAlaAsp pol  
 AAGAAATATCCGAGATTCCGAATGCCACAGGATTTGGATCTGATAACGGGCTGCTTCACTCCAGGTAACTGCTGGTGGCCG  
 5140 5150 5160 5170 5180 5190 5200 5210 5220

LeuLeuGlyIleAspTrpLysLeuHisCysAlaTyrArgProGlnSerSerGlyGlnValGluArgMetAsnArgThrIleLysGluThr pol  
 ATTTAGCTGGGATCGATTGAAATTAACATGCTGCTATAGACCCAGACTTCAGTTCAGGTAGAAAATGAATAGAACCTCAAGGAGA  
 5230 5240 5250 5260 5270 5280 5290 5300 5310

LeuThrLysLeuThrLeuAlaAlaGlyThrArgAspTrpValLeuLeuProLeuAlaLeuTyrArgAlaArgAsnThrProGlyPro pol  
 CTCTAACTAAATTAACGCTTGCAGCTGACAGACTAGAGACTGGTACTCTACTCCCTTACTGCTTTACCGAGCCCGCAACTCCGGCC  
 5320 5330 5340 5350 5360 5370 5380 5390 5400

HisGlyLeuThrProTyrGluIleLeuTyrGlyAlaProProLeuValAsnPheHisAspProAspMetSerGluLeuThrAsnSer pol  
 CCCATGGACTGACTCGTATGAAATCTGTACGGGGCCCGCCCTGTTAACTTCCATGACCCGCAATCTGAGAACTTAATAATA  
 5410 5420 5430 5440 5450 5460 5470 5480 5490

ProSerLeuGlnAlaHisLeuGlnAlaLeuGlnThrValGlnArgGluIleTrpLysProLeuAlaGluAlaTyrArgAspGlnLeuAsp pol  
 GCCCATCTCTCGAAGCTCACTAGCCCTCCAAAGCGGTGACGCAAAATTTGAAAGCACTGGCCGAGCCTACCGGACCAACTAGC  
 5500 -----Oligo. 5520 5530 5540 5550 5560 5570 5580

GlnProValIleProHisProPheArgIleGlyAspSerValTrpValArgArgHisGlnThrLysAsnLeuGluProArgTrpLysGly pol  
 ACCAACCTGATACCCACCCCTTCCGATTGGAGACTCCGCTGGCTCCCGGCAAGCAGCAAACTTAGAACTCGCTGAGG  
 5590 5600 5610 5620 5630 5640 5650 5660 5670

ProTyrThrValLeuLeuThrThrProThrAlaLeuValAspGlyIleSerAlaTrpIleHisAlaHisValLysAlaAlaThr pol  
 GACCTTACCGCTCACTAGCCCTCCAAAGCGGTGACGCAAAATTTGAAAGCACTGGCCGAGCCTACCGGACCAACTAGC  
 5680 5690 5700 5710 5720 5730 5740 5750 5760

FIG. 1—Continued

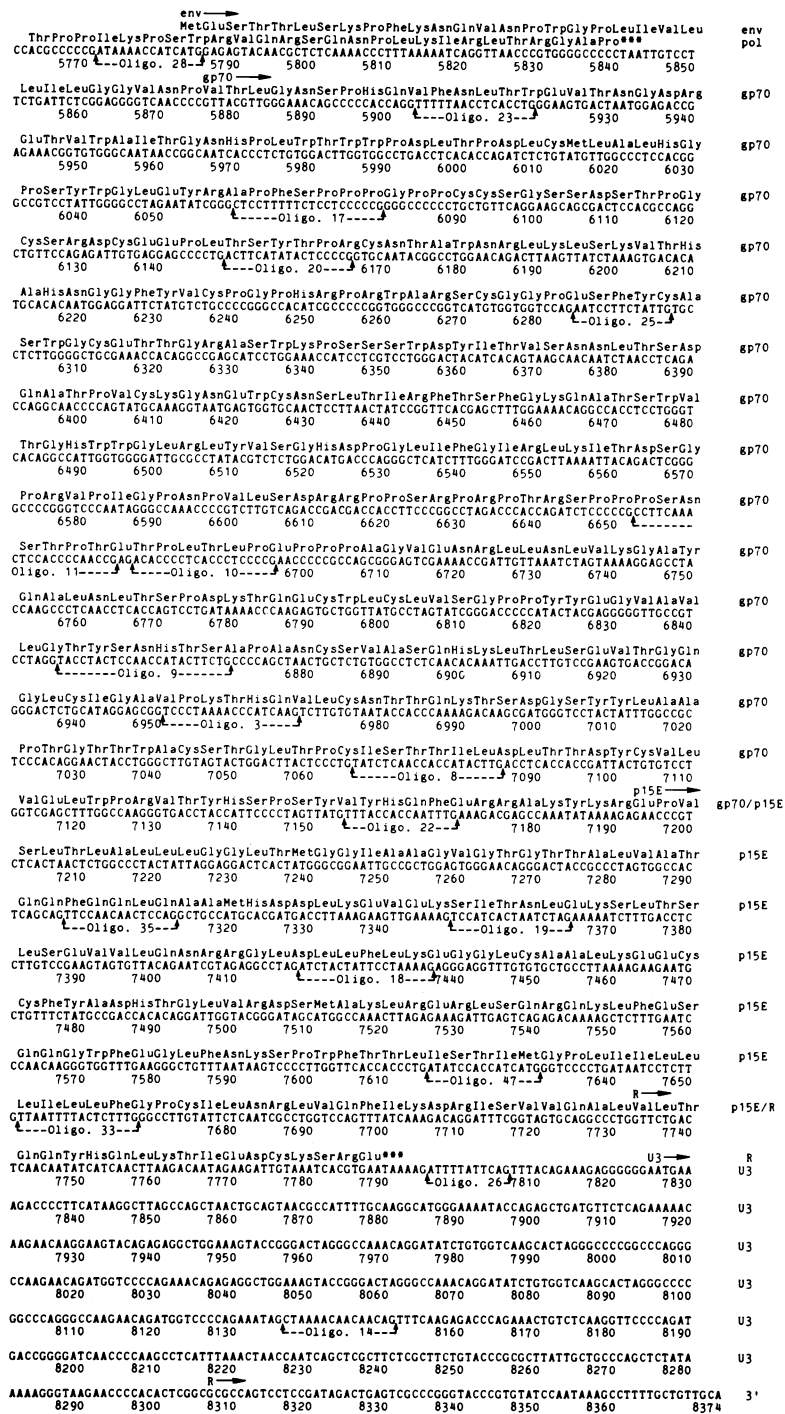


FIG. 1. Nucleotide sequence of AKV MLV. The DNA strand corresponding to the viral plus strand is shown. The amino acid sequences of the *gag*, *pol*, and *env* genes, as well as the beginning of the R, U5, and U3 sequences, are noted above the sequence. RNase T<sub>1</sub> oligonucleotides (23) and features discussed in the text are indicated below the sequence.

identify any differences from the sequences of this region previously reported (15, 21, 56).

The AKV and Moloney MLV (Mo-MLV) (48) nucleotide sequences were compared by using the DIAGON computer program (52).

## RESULTS AND DISCUSSION

Figure 1 shows the 8,374-nucleotide sequence corresponding to the plus (sense) strand of the AKV RNA genome, as determined from the infectious molecular clone of AKV, AKR-623 (22). Because the AKV terminal repeat (R) is nearly identical to that of Mo-MLV (56), the 5' and 3' termini of AKV have been aligned to match the Mo-MLV genome (48). Figure 1 also shows the deduced amino acid sequence of the AKV *gag*, *pol*, and *env* gene products, the limits of the R sequence, and the locations of the unique 5' (U5) and 3' (U3) sequences that are duplicated in the provirus LTR. These structural features were identified by comparing the AKV sequence with the nucleotide sequence of Mo-MLV (48) and the amino acid sequences of Rauscher MLV proteins determined by Orozlan and his colleagues (see reference 58).

**Comparison with previously reported AKV sequences.** The nucleotide sequence shown in Fig. 1 of the AKR-623 LTR (nucleotides 1 to 144 and 7,825 to 8,374) is identical to that reported by Van Beveren et al. (56). Lenz et al. (21) have reported the nucleotide sequence of the AKR-623 *env* gene (nucleotides 5,674 to 7,865). The sequence shown in Fig. 1 differs at two positions: an additional T residue at nucleotide 5,676 and a C rather than G residue at nucleotide 5,740.

We previously reported the sequence of nucleotides 3,309 through 7,865 from a noninfectious molecular clone of AKV (15). There are six differences within this region between the infectious and noninfectious clones: the noninfectious clone contains an extra G residue within the polypurine tract near the origin of plus-strand strong stop synthesis (nucleotides 7,819 to 7,824) and contains an A residue in place of a G residue at five positions (nucleotides 4,685, 5,240, 5,885, 7,169, and 7,556). Curiously, all five of the base substitutions are identical. Furthermore, the affected guanine residues are surrounded by the consensus sequence TTNGAAA (the consensus nucleotides are each present in at least four of the five sites), suggesting that this sequence is prone to mutation during viral replication. Consistent with this, two differences between the LTRs of AKV and Gross MLV (57) and a change between the 5'-leader sequences of two different molecular clones of Mo-MLV (47) are also G-to-A transitions which share this consensus sequence. The G-to-A transition at nucleotide 5,240 of AKV has mutated a tryptophan codon (UGG) into a UGA termination codon within the *pol* gene of the noninfectious clone, perhaps explaining the lack of infectivity.

Because the 3' half of AKV has already been described extensively (15, 21, 56), these sequences will not be discussed further here.

**AKV 5'-leader sequence.** The 5'-leader region of AKV stretches from the 5' cap site to the first initiation codon (AUG) in the AKV viral genome, at nucleotides 639 to 641, where translation of the *gag* and *pol* precursors is initiated. Within the 5' leader of AKV, the 5'-terminal 220 nucleotides are 90% homologous to the analogous sequences in Mo-MLV (48). This stretch of nucleotides contains the R (nucleotides 1 to 68) and U5 (nucleotides 69 to 144) sequences and the tRNA<sup>pro</sup> binding site (nucleotides 145 to 162) where negative-strand synthesis initiates, as well as a putative

donor splice site sequence (nucleotides 203 to 209) for generation of the *env* mRNA. A possible *env* mRNA acceptor splice site sequence, identical to Mo-MLV, lies at nucleotides 5,508 to 5,516.

The region of the AKV genome spanning nucleotides 220 to 270 is characterized by a stretch of 21 alternating purine/pyrimidine residues (nucleotides 245 to 265). This sequence, which is almost entirely composed of repeating TG dinucleotides, is remarkable for three reasons: (i) repeating tracts of TG dinucleotides can exist as a left-handed Z helix under physiological conditions (13, 30); (ii) this sequence is found highly repeated in eucaryotic genomes (11, 25); and (iii) the sequence has been implicated as a hot spot for recombination (35, 49, 54). The analogous region of the Mo-MLV genome bears little resemblance to AKV and lacks the stretch of TG residues, indicating that the latter is not essential for replication. Nevertheless, Mo-MLV and AKV do share within this region two copies of the pentanucleotide TGTCT (nucleotides 227 to 231 and 239 to 243 in AKV), which is also frequently present neighboring other TG tracts identified in mammalian genomes (12, 25, 27, 28). These results suggest that this region of both AKV and Mo-MLV may be related to the TG tract sequences found in eucaryotic genomes.

The remainder of the AKV leader sequence is characterized by an open reading frame beginning at nucleotide 318 in the same frame as the *gag* coding sequences. An analogous open reading frame begins at nucleotide 310 of the Mo-MLV genome. (The two open reading frames are out of frame with respect to one another until nucleotide 363 in the AKV genome.) Cells infected by MLV express on their surface a glycosylated *gag* gene product, referred to as gPr80<sup>gag</sup> (20, 50, 55). Peptide mapping of gPr80<sup>gag</sup> (8, 45) indicated that the amino-terminal region of gPr80<sup>gag</sup> probably derives from part of the open reading frame found in the 5'-leader region. Schwartzberg et al. (47) have shown that Mo-MLV deletion mutants which lack parts of this open reading frame do not produce gPr80<sup>gag</sup> but are still replication competent. It is perhaps not surprising, therefore, that the AKR-623 AKV clone is infectious even though it lacks an AUG initiation codon to express gPr80<sup>gag</sup>.

**AKV *gag* and *pol* genes.** The *gag* and *pol* gene products are encoded by a long open reading frame (nucleotides 639 to 5,843) that is interrupted by a single amber (UAG) codon separating the two genes. The MLV *gag* gene is translated as a single polypeptide, Pr65<sup>gag</sup>, which is subsequently cleaved to produce the internal viral proteins p15 (encoded by nucleotides 642 to 1,025), p12 (1,026 to 1,280), p30 (1,281 to 2,069), and p10 (2,070 to 2,237). Comparison of the AKV sequence in this region with the Mo-MLV and Rauscher MLV *gag* sequences (48, 58) shows that the amino acid residues adjacent to the proteolytic cleavage sites are identical in all three MLVs. The AKV p30 and p10 peptides are very highly conserved with respect to Mo-MLV and Rauscher MLV (95% homologous at the amino acid level), whereas the p12 and p15 peptides are not.

The initial translation product of the *pol* gene is the Pr180<sup>gag-pol</sup> fusion protein that contains both *gag* and *pol* antigenic determinants. To synthesize such a molecule, the amber codon separating *gag* and *pol* in the viral genome must be either eliminated or suppressed. This could be accomplished by either splicing the genomic RNA to remove the UAG codon from the *pol* mRNA, as appears to be the case with Rous sarcoma virus (46), or translation of the termination codon by a suppressor tRNA (26, 34).

Shinnick et al. (48) proposed a RNA secondary structure

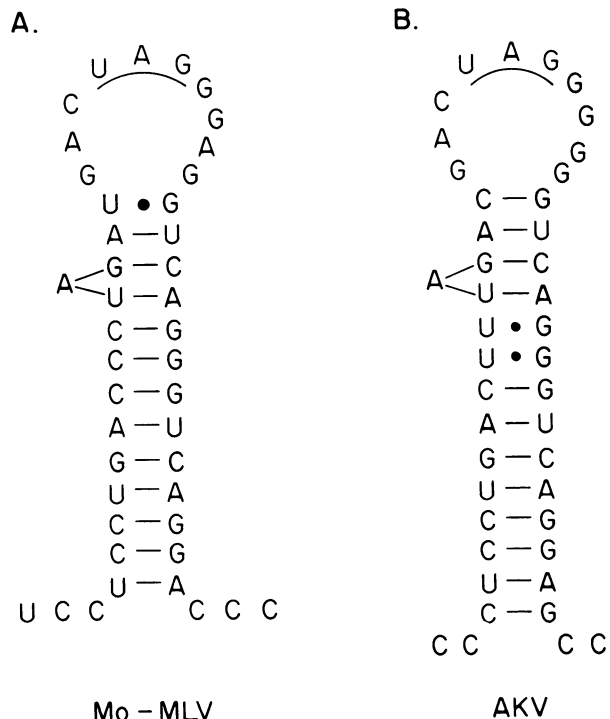


FIG. 2. Comparison of potential RNA secondary structures flanking the amber codon (underlined) which separates the *gag* and *pol* genes of Mo-MLV (A) and AKV (B).

flanking the amber codon separating the *gag* and *pol* genes in Mo-MLV. This and an analogous secondary structure from the AKV sequence are shown in Fig. 2. The nucleotides in AKV that are base paired in the secondary structure are marked by apostrophes below the sequence shown in Fig. 1

(between nucleotides 2,232 and 2,270). Although there are several base differences between the Mo-MLV and AKV structures (one of which extends the AKV stem by 1 base pair), none of these destroy any base pairs, providing further evidence that such a structure exists *in vivo*. The stem of each structure contains a single, unpaired, "bulged" adenine residue. Double-stranded RNA helices containing a bulged nucleotide have been implicated as RNA-protein contact sites (32); the stem-and-loop structure shown in Fig. 2 might thus serve as a binding site for a regulatory protein that controls the relative amounts of Pr65<sup>gag</sup> and Pr180<sup>gag-pol</sup>, whether this occurs by splicing of the genomic RNA or tRNA suppression.

The *pol* gene in AKV extends from nucleotide 2,253 to 5,843 and can encode 1,196 amino acids. As noted previously in Mo-MLV (48), this reading frame is much larger than required to encode the 80,000-dalton reverse transcriptase polypeptide. Because the *pol* gene of avian retroviruses encodes a double-stranded DNA endonuclease (10, 44), Shinnick et al. (48) suggested that the extra coding potential of the MLV *pol* gene might encode an analogous endonuclease; recent evidence (T. Shinnick, personal communication) indicates that an endonuclease activity is in fact encoded by the 3' region of the *pol* gene in Mo-MLV. Comparison of the AKV and Mo-MLV *pol* amino acid sequences shows that the amino-terminal two-thirds (encoded by nucleotides 2,253 to 4,500) of the *pol* gene product is 96% homologous, whereas the carboxy-terminal third is only 77% homologous. The conserved region is the expected size to encode reverse transcriptase, suggesting that it is the endonuclease activity which is not highly conserved and might, therefore, be virus specific.

**Structure of oncogenic MCF MLV.** The structure of recombinant MCF viruses has been extensively studied by both restriction site mapping of proviral DNA (4) and RNase T<sub>1</sub> oligonucleotide mapping (23, 38). Using the RNase T<sub>1</sub> oligonucleotide sequences determined by Pedersen and Haseltine

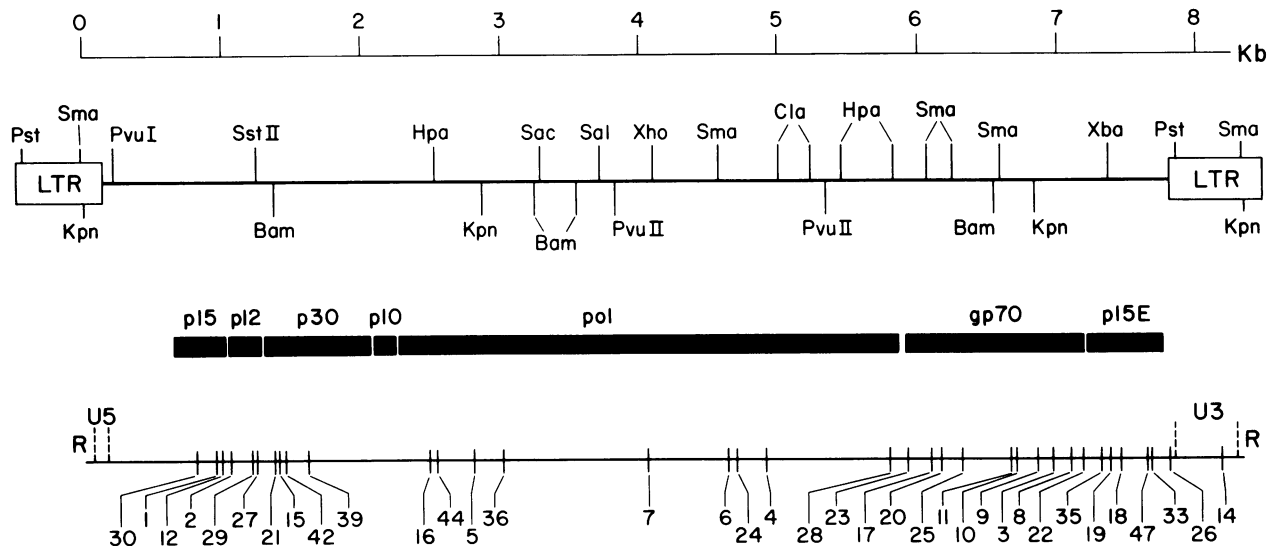


FIG. 3. Correlation of AKV restriction sites and RNase T<sub>1</sub> oligonucleotides with the AKV gene map. Upper portion shows the location of the provirus restriction enzyme sites mapped by Chattopadhyay et al. (4) and confirmed by the sequence. Unless otherwise specified all enzymes are I. Wild-type AKV contains a *Hind*III site approximately 2.5 kilobases from the 5' cap site (53) that is missing from the AKR-623 clone. AKV does not contain the recognition sites for the following enzymes: *Avai*III, *Bcl*I, *Eco*RI, *Mlu*I, *Mst*I, *Nde*I, *Nru*I, *Sph*I, and *Xma*III. Middle portion shows the AKV gene map. At the bottom is shown the location of the RNase T<sub>1</sub> oligonucleotides used by Lung et al. (23) to study the structure of MCF genomes. Oligonucleotide 39 is the N-tropic specific oligonucleotide identified by Rommelaere et al. (37).

(33) and the pancreatic RNase digestion products identified by Rommelaere et al. (38), we have located within the AKV sequence the RNase T<sub>1</sub> oligonucleotides used by Lung et al. (23) to study the recombination patterns of MCF viruses. These RNase T<sub>1</sub> oligonucleotides are identified under the sequence in Fig. 1. Figure 3 correlates the location of the RNase T<sub>1</sub> oligonucleotides (bottom) with both the AKV provirus restriction sites (top) and the AKV gene map (middle).

The sequence from the 5' half of AKV shows that all of the RNase T<sub>1</sub> oligonucleotides analyzed by Lung et al. (23) from this region derive from either the p15, p12, and amino-terminal half of the p30 coding sequences or a short region in *pol*. The exact nature of large regions of the 5' half of MCF genomes were thus ignored in the oligonucleotide studies. Furthermore, because the restriction site map of the 5' half of AKV is more similar to that of nonectropic proviruses than is the AKV 3' half (4), the restriction site analyses of Chattopadhyay et al. (4) were likewise not able to detect recombination as readily within the 5' half of the MCF genomes as within the 3' half.

Nevertheless, Lung et al. (23) showed that many of the oncogenic MCF viruses (e.g., MCF 13) lack specific RNase T<sub>1</sub> oligonucleotides found in AKV, for example, the *pol* oligonucleotides 44 and 5 (see Fig. 3). Consistent with this result Chattopadhyay et al. (4) showed that some of the MCF proviruses (such as MCF 13) lack AKV restriction sites (e.g., *SacI* and *HpaI*) which lie within the *pol* gene. Taken together, these results suggest that sequences within the 5' half of MCF viruses are frequently derived from the non-AKV parent virus. The new 5' AKV sequences presented here should aid in identifying the exact locations of the nonectropic sequences within the 5' region of MCF genomes.

#### ACKNOWLEDGMENTS

This work was done in the laboratory of F. Sanger, whom I thank for support and encouragement. I thank D. Lowy and S. Chattopadhyay for the AKR-623 clone of AKV; A. Bankier, B. Barrell, M. Biggin, and T. Gibson for advice on dideoxy sequencing; C. Keller, R. Roberts, and R. Staden for help with the computer analyses; J. Abraham, S. Hughes, and J. Sambrook for critical readings of the manuscript; and P. Barkley for typing the manuscript.

I was supported by a fellowship from the Helen Hay Whitney Foundation.

#### LITERATURE CITED

- Anderson, S. 1981. Shotgun DNA sequencing using cloned DNaseI-generated fragments. *Nucleic Acids Res.* **9**:3015-3027.
- Biggin, M. D., T. J. Gibson, and G. F. Hong. 1983. Buffer gradient gels and <sup>35</sup>S label as an aid to rapid DNA sequence determination. *Proc. Natl. Acad. Sci. U.S.A.* **80**:3963-3965.
- Chattopadhyay, S. K., M. W. Cloyd, D. L. Linemeyer, M. R. Lander, E. Rands, and D. R. Lowy. 1982. Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. *Nature (London)* **295**:25-31.
- Chattopadhyay, S. K., M. R. Lander, S. Gupta, E. Rands, and D. R. Lowy. 1981. Origin of mink cytopathic focus-forming (MCF) viruses: comparison with ecotropic and xenotropic murine leukemia virus genomes. *Virology* **113**:465-483.
- Chien, Y.-H., I. M. Verma, T. Y. Shih, E. M. Scolnick, and N. Davidson. 1978. Heteroduplex analysis of the sequence relations between the RNAs of mink cell focus-inducing and murine leukemia viruses. *J. Virol.* **28**:352-360.
- Cloyd, M. W., J. W. Hartley, and W. P. Rowe. 1980. Lymphomagenicity of recombinant mink cell focus-inducing murine leukemia viruses. *J. Exp. Med.* **151**:542-552.
- Deininger, P. L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* **129**:216-223.
- Edwards, S. A., and H. Fan. 1980. Sequence relationship of glycosylated and unglycosylated *gag* polyproteins of Moloney murine leukemia virus. *J. Virol.* **35**:41-51.
- Elder, J. H., J. W. Gautsch, F. C. Jensen, R. A. Lerner, J. W. Hartley, and W. P. Rowe. 1977. Biochemical evidence that MCF murine leukemia viruses are envelope (*env*) gene recombinants. *Proc. Natl. Acad. Sci. U.S.A.* **74**:4676-4680.
- Grandgenett, D. P., A. C. Vora, and R. D. Schiff. 1978. A 32,000-dalton nucleic acid-binding protein from avian retrovirus cores possesses DNA endonuclease activity. *Virology* **89**:119-132.
- Hamada, H., M. G. Petrino, and T. Kakunaga. 1982. A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proc. Natl. Acad. Sci. U.S.A.* **79**:6465-6469.
- Hamada, H., M. G. Petrino, and T. Kakunaga. 1982. Molecular structure and evolutionary origin of human cardiac muscle actin gene. *Proc. Natl. Acad. Sci. U.S.A.* **79**:5901-5905.
- Haniford, D. B., and D. E. Pulleyblank. 1983. Facile transition of poly [d(TG) · d(CA)] into a left-handed helix in physiological conditions. *Nature (London)* **302**:632-634.
- 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.
- Herr, W., V. Corbin, and W. Gilbert. 1982. Nucleotide sequence of the 3' half of AKV. *Nucleic Acids Res.* **10**:6931-6944.
- Herr, W., and W. Gilbert. 1983. Somatic acquired recombinant murine leukemia proviruses in thymic leukemias of AKR/J mice. *J. Virol.* **46**:70-82.
- Herr, W., D. Schwartz, and W. Gilbert. 1983. Isolation and mapping of cDNA probes specific for ecotropic and nonectropic murine leukemia proviruses. *Virology* **125**:139-154.
- Kaplan, H. S. 1967. On the natural history of murine leukemias: presidential address. *Cancer Res.* **27**:1325-1340.
- Kelly, M., C. A. Holland, M. L. Lung, S. K. Chattopadhyay, D. R. Lowy, and N. H. Hopkins. 1983. Nucleotide sequence of the 3' end of MCF 247 murine leukemia virus. *J. Virol.* **45**:291-298.
- Ledbetter, J., and R. C. Nowinski. 1977. Identification of the Gross cell surface antigen associated with murine leukemia virus-infected cells. *J. Virol.* **23**:315-322.
- Lenz, J., R. Crowther, A. Straceski, and W. Haseltine. 1982. Nucleotide sequence of the AKV *env* gene. *J. Virol.* **42**:519-529.
- Lowy, D. R., E. Rands, S. K. Chattopadhyay, C. F. Garon, and G. L. Hager. 1980. Molecular cloning of infectious integrated murine leukemia virus DNA from infected mouse cells. *Proc. Natl. Acad. Sci. U.S.A.* **77**:614-618.
- Lung, M. L., J. W. Hartley, W. P. Rowe, and N. H. Hopkins. 1983. Large RNase T<sub>1</sub>-resistant oligonucleotides encoding p15E and the U3 region of the long terminal repeat distinguish two biological classes of mink cell focus-forming type C viruses of inbred mice. *J. Virol.* **45**:275-290.
- Messing, J., and J. Vieira. 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. *Gene* **19**:269-276.
- Miesfeld, R., M. Krystal, and N. Arnheim. 1981. A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human δ and β globin genes. *Nucleic Acids Res.* **9**:5931-5947.
- Murphy, E. C., N. Wills, and R. B. Arlinghaus. 1980. Suppression of murine retrovirus polypeptide termination: effect of amber suppressor tRNA on the cell-free translation of Rauscher murine leukemia virus, Moloney murine leukemia virus, and Moloney murine sarcoma virus 124 RNA. *J. Virol.* **34**:464-473.
- 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.
- Nishioka, Y., and P. Leder. 1980. Organization and complete

- sequence of identical embryonic and plasmacytoma  $\kappa$  V-region genes. *J. Biol. Chem.* **255**:3691–3694.
29. Nishizuka, Y., and K. Nakakuki. 1968. Acceleration of leukemogenesis in AKR mice by grafts, cell suspensions, and cell-free centrifugates of thymuses from preleukemic AKR donors. *Int. J. Cancer* **3**:203–210.
  30. Nordheim, A., and A. Rich. 1983. The sequence  $(dC-dA)_n \cdot (dG-dT)_n$  forms left-handed Z-DNA in negatively supercoiled plasmids. *Proc. Natl. Acad. Sci. U.S.A.* **80**:1821–1825.
  31. O'Donnell, P. V., E. Stockert, Y. Obata, and L. J. Old. 1981. Leukemogenic properties of AKR dultropic (MCF) viruses: amplification of murine leukemia virus-related antigens on thymocytes and acceleration of leukemia development in AKR mice. *Virology* **112**:548–563.
  32. Peattie, D. A., S. Douthwaite, R. A. Garrett, and H. F. Noller. 1981. A "bulged" double helix in a RNA-protein contact site. *Proc. Natl. Acad. Sci. U.S.A.* **78**:7331–7335.
  33. Pedersen, F. S., and W. A. Haseltine. 1980. Analysis of the genome of an endogenous, ecotropic retrovirus of the AKR strain of mice: micromethod for detailed characterization of high-molecular-weight RNA. *J. Virol.* **33**:349–365.
  34. Philipson, L., P. Andersson, U. Olshevsky, R. Weinberg, D. Baltimore, and R. Gesteland. 1978. Translation of MuLV and MSV RNAs in nuclease-treated reticulocyte extracts: enhancement of the *gag-pol* polypeptide with yeast suppressor tRNA. *Cell* **13**:189–199.
  35. Proudfoot, N. J., and T. Maniatis. 1980. The structure of a human  $\alpha$ -globin pseudogene and its relationship to  $\alpha$ -globin gene duplication. *Cell* **21**:537–544.
  36. Quint, W., W. Quax, H. van der Putten, and A. Berns. 1981. Characterization of AKR murine leukemia virus sequences in AKR mouse substrains and structure of integrated recombinant genomes in tumor tissues. *J. Virol.* **39**:1–10.
  37. Rommelaere, J., H. Donis-Keller, and N. Hopkins. 1979. RNA sequencing provides evidence for allelism of determinants of the N-, B- or NB-tropism of murine leukemia viruses. *Cell* **16**:43–50.
  38. Rommelaere, J., D. V. Faller, and N. Hopkins. 1978. Characterization and mapping of RNase T<sub>1</sub>-resistant oligonucleotides derived from the genomes of AKV and MCF murine leukemia viruses. *Proc. Natl. Acad. Sci. U.S.A.* **75**:495–499.
  39. Rowe, W. P. 1972. Studies of genetic transmission of murine leukemia virus by AKR mice. *J. Exp. Med.* **136**:1272–1285.
  40. Rowe, W. P. 1978. Leukemia virus genomes in the chromosomal DNA of the mouse. *Harvey Lect.* **71**:173–192.
  41. Rowe, W. P., and T. Pincus. 1972. Quantitative studies of naturally occurring murine leukemia virus infection of AKR mice. *J. Exp. Med.* **135**:429–436.
  42. Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe. 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. *J. Mol. Biol.* **143**:161–178.
  43. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* **74**:5463–5467.
  44. Schiff, R. D., and D. P. Grandgenett. 1978. Virus-coded origin of a 32,000-dalton protein from avian retrovirus cores: structural relatedness of p32 and the  $\beta$  polypeptide of the avian retrovirus DNA polymerase. *J. Virol.* **28**:279–291.
  45. Schultz, A. M., and S. Orozlan. 1978. Murine leukemia virus *gag* polyproteins: the peptide chain unique to Pr80 is located at the amino terminus. *Virology* **91**:481–486.
  46. Schwartz, D. E., R. Tizard, and W. Gilbert. 1983. Nucleotide sequence of Rous sarcoma virus. *Cell* **32**:853–869.
  47. Schwartzberg, P., J. Colicelli, and S. P. Goff. 1983. Deletion mutants of Moloney leukemia virus which lack glycosylated *gag* protein are replication competent. *J. Virol.* **46**:538–546.
  48. Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukemia virus. *Nature (London)* **293**:543–548.
  49. Slightom, J. L., A. E. Blechl, and O. Smithies. 1980. Human fetal G $\gamma$ - and A $\gamma$ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**:627–638.
  50. Snyder, H. W., E. Stockert, and E. Fleissner. 1977. Characterization of molecular species carrying Gross cell surface antigen. *J. Virol.* **23**:302–314.
  51. Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. *Nucleic Acids Res.* **10**:4731–4751.
  52. Staden, R. 1982. An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Res.* **10**:2951–2961.
  53. Steffen, D. L., S. Bird, and R. A. Weinberg. 1980. Evidence for the asiatic origin of endogenous AKR-type murine leukemia proviruses. *J. Virol.* **35**:824–835.
  54. Stringer, J. R. 1982. DNA sequence homology and chromosomal deletion at a site of SV40 DNA integration. *Nature (London)* **296**:363–366.
  55. Tung, J.-S., T. Yoshiki, and E. Fleissner. 1976. A core polyprotein of murine leukemia virus on the surface of mouse leukemia cells. *Cell* **9**:573–578.
  56. 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.
  57. Villemur, R., E. Rassart, L. DesGroseillers, and P. Jolicouer. 1983. Molecular cloning of viral DNA from leukemogenic Gross Passage A murine leukemia virus and nucleotide sequence of its long terminal repeat. *J. Virol.* **45**:539–546.
  58. Weiss, R., N. Teich, H. Varmus, and J. Coffin. 1982. RNA tumour viruses, p. 1378–1382. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.