

Heteroduplex Analysis of the Sequence Relations Between the RNAs of Mink Cell Focus-Inducing and Murine Leukemia Viruses†

YUEH-HSIU CHIEN,¹ INDER M. VERMA,² THOMAS Y. SHIH,³ EDWARD M. SCOLNICK,³ AND NORMAN DAVIDSON^{1*}

Department of Chemistry, California Institute of Technology, Pasadena, California 91125,¹ Tumor Virology Laboratory, The Salk Institute, San Diego, California 92112,² and Laboratory of Tumor Virus Genetics, National Cancer Institute, Bethesda, Maryland 20014³

Received for publication 1 May 1978

The sequence relationships between AKR ecotropic virus and an AKR-derived "mink cell focus-inducing" (MCF) isolate (AKR MCF 247), between Moloney murine leukemia virus (M-MLV) and an M-MLV MCF isolate (M-MLV₈₃), and between AKR and M-MLV were studied by electron microscopic heteroduplex analysis. The MCF-specific sequences were found to map from 1.95 kilobases (kb) to 2.75 kb (\pm 0.15 kb) from the 3' end of the RNAs for both MCF isolates. The major sequence nonhomology regions between AKR and M-MLV lie between 0.9 and 3.5 kb from the 3' end. However, the AKR and M-MLV sequences immediately adjacent to the 1.95- and 2.75-kb junctions with MCF-specific sequences are relatively similar in AKR and M-MLV. Our results suggest that the *env* gene of MLVs maps from 1 kb to 3 kb from the 3' end of the genomic RNA and that the carboxyl end of the glycoprotein of each MCF strain is similar (or identical) to that of its ecotropic parent.

Recently, a novel class of murine leukemia virus (MLV) associated with the development of spontaneous lymphomas has been described (10). These viruses were detected in the thymuses of leukemic and late preleukemic AKR mice, preleukemic C58 mice, and in the lymphomas developing in NIH Swiss mice carrying the AKR ecotropic virus-inducing loci AKV-1 or AKV-2 (10). The viruses induce focal areas of morphological alteration in a mink lung fibroblast cell line and hence are referred to as "mink cell focus-inducing" (MCF) strains (10). They have the dual host range of both xenotropic and N-tropic ecotropic MLVs. Hartley et al. proposed that MCF strains arise *in vivo* by recombination and are recombinants between ecotropic AKR virus and endogenous xenotropic MLVs (10). Two types of biochemical experiments support this notion. Tryptic peptide analyses indicate that the envelope glycoprotein, gp70, of several AKR MCF isolates contains peptides in common with both ecotropic and xenotropic gp70's, but the other virion proteins of the MCF strains are indistinguishable from those of AKR virus (4). Furthermore, some peptides which are specific for the gp70's of AKR or xenotropic viruses are missing from the MCF

gp70's peptide analyses, suggesting that the MCF gp70's are not simply mixtures of AKR and xenotropic gp70's. Rommelaere et al. have shown that 55 to 75% of the T1-resistant oligonucleotides of AKV-1 and AKV-2 RNAs are present in the RNAs of several MCF virus strains isolated from AKR mice or from NIH Swiss mice carrying the AKV-inducing locus. The majority of the oligonucleotides unique to MCF viruses are clustered and lie in the 3' half of the genome. Since the *env* gene is believed to be located in the 3' half of the MLV genome, it appears that the difference between genomes of the AKV-1 and AKV-2 and the MCF RNAs lie in the *env* gene (19).

Similarly, Shih et al. (21) have isolated a strain of Moloney MLV (M-MLV₈₃) which demonstrates a broadened host range, allowing growth on both mouse cells and cells permissive for xenotropic viruses. Furthermore, upon growth on mink lung fibroblasts it induces cytopathic foci diagnostic of MCF virus. Fingerprint analyses of the genomes of M-MLV₈₃ and parental M-MLV show a cluster of about seven large T1 oligonucleotides unique to M-MLV₈₃ and about 10 unique to M-MLV. These unique oligonucleotides were mapped between 2,000 and 4,000 nucleotides from the 3' end of the genome. All these data suggest that, like AKR MCF, M-

† Contribution no. 5769 from the Department of Chemistry, California Institute of Technology.

MLV₈₃ is also a recombinant virus. The recombination appears to occur in the *env* gene.

The HIX virus previously isolated by Fischinger et al. (8) is similar to the MCF viruses in that it appears to be a recombinant with both M-MLV and xenotropic determinants in its *env* protein.

To obtain further information about possible regions for recombination, and about regions possibly playing a role in the focus-forming property of the MCF strains, we have studied the sequence relations of two particular isolates, AKR MCF 247 and M-MLV₈₃, to the corresponding AKR and M-MLV viral genomes, by electron microscopic heteroduplex analysis. It has been reported that M-MLV and AKR viruses have different envelope glycoproteins (5, 13). Therefore, we also studied the sequence relationship between the AKR and M-MLV genomes.

MATERIALS AND METHODS

Viruses. A cloned isolate of M-MLV (clone 1 M-MLV) was grown in NIH 3T3 cells, and virus was harvested as described before (7).

AKR ecotropic virus was obtained from three different sources: (i) AKR ecotropic virus grown in fibroblasts from the Office of Program Resources and Logistics, National Cancer Institute; (ii) AKV-2 virus grown in wild-type mouse embryo fibroblast cell line (SC-1); and (iii) from S. Chattopadhyaya of the National Institutes of Health. By heteroduplex analysis, the RNAs obtained from these sources were indistinguishable. AKR virus obtained from the National Cancer Institute was used for complementary DNA (cDNA) synthesis.

SC-1 cells infected with AKR MCF 247 virus were obtained from Fred Jensen of Scripps Clinic and Research Foundation, La Jolla, Calif. Virus was harvested at 3-h intervals and purified as described (25).

M-MLV clone 83 (M-MLV₈₃) was isolated by end-point dilution cloning in Microtest II culture plates on SC-1 cells from a stock of M-MLV originally obtained from R. Bassin of the National Cancer Institute (23). It was propagated on a mink lung fibroblast cell line.

Unlabeled deoxyribonucleoside triphosphates were obtained from P-L Biochemicals, and [³H]dTTP was obtained from Schwartz/Mann.

Preparation of RNA. The method used for preparing 60S to 70S viral RNA from purified virions has been described (6). RNA from peak fractions was pooled and precipitated with 2.5 volumes of ethanol. The precipitate was dissolved in 2.5 mM EDTA, pH 7, and heated to 80°C for 1 min. After adjusting the NaCl concentration to 0.5 M, the solution was passed through an oligodeoxythymidylic acid-cellulose column (T-3, Collaborative Research) to select polyadenylic acid [poly(A)]-containing molecules.

Synthesis of cDNA. Genome-length cDNA transcripts were synthesized by use of purified virions of AKR and M-MLV as described previously (25).

Briefly, the reaction mixture contained 50 mM Tris-hydrochloride, pH 8.3, 10 mM dithiothreitol, 6 mM magnesium acetate, 60 mM NaCl, 2 mM each of dATP, dCTP, dGTP, and [³H]dTTP (specific activity, 40 cpm/pmol). About 2 to 2.5 mg of purified virions was used in 1.0-ml reaction mixtures. The virions were lysed with 0.01% Nonidet P-40 and incubated for 8 to 18 h at 37°C. The nucleic acids were precipitated with ethanol after extraction and deproteinization with phenol and chloroform. Samples were analyzed by velocity sedimentation on a 5 to 20% alkaline sucrose gradient (0.3 M KOH, 0.7 M LiCl, 0.005 M EDTA, pH 12.4). Material sedimenting faster than form II polyoma DNA marker (18S) was collected, neutralized, and used in the experiments described here. The lengths of these molecules as determined by electron microscopy were distributed approximately uniformly between 6 and 10 kilobases (kb) (data not shown).

Hybridization. RNA:cDNA hybridization was carried out as follows. A mixture of 1.5 µg of 35S RNA/ml and 1 µg of cDNA/ml in 80% formamide, 0.4 M NaCl, 0.01 M piperazine-*N,N'*-bis(2-ethanesulfonic acid), and 0.001 M EDTA, pH 6.3 was incubated at 50°C for 30 min. Samples were incubated with simian virus 40 (SV40)-polydeoxybromouridylic acid [poly(dBrU)] for mapping the poly(A) ends of the RNA (2). [We now use SV40-poly(dBrU) instead of SV40-polydeoxythymidylic acid as a poly(A) label, because of the greater stability of poly(dBrU):poly(A) duplexes.] The solution was adjusted to a final concentration of 60% formamide for the AKR/AKR MCF 247 and M-MLV/M-MLV₈₃ heteroduplexes and to 60% urea-formamide (16) for the AKR/M-MLV heteroduplexes and was spread as described (3).

The circular SV40 molecules present in the reaction also served as an internal length standard. Single-stranded φX174 molecules (5.37 kb) and double-stranded SV40 molecules (5.27 kb) were spread together in separate experiments to establish a double-stranded to single-stranded length ratio for an equal number of nucleotides of 1.32. The coordinates of various sequence homology and nonhomology regions were measured from the labeled 3' end of the RNA. The length of SV40 molecules is taken as 1.76 µm.

RESULTS

Molecular lengths of RNAs from AKR, AKR MCF 247, M-MLV, and M-MLV₈₃ were measured by agarose gel electrophoresis in the presence of 5 mM CH₃HgOH as previously described (1), except that the concentration of EDTA in the electrophoresis buffer was reduced from 1 mM to 0.1 mM. Each viral RNA showed a single band on the gel by ethidium bromide staining. By using Sindbis viral RNA (13 kb), mouse 28S rRNA (5.3 kb), 18S rRNA (2.06 kb), *Escherichia coli* 23S rRNA (3.2 kb), and 16S rRNA (1.64 kb) as markers, the molecular length of each of the viral RNAs was measured as 9.7 ± 0.5 kb (data not shown).

AKR/AKR MCF 247 heteroduplex. A typical heteroduplex formed between AKR cDNA

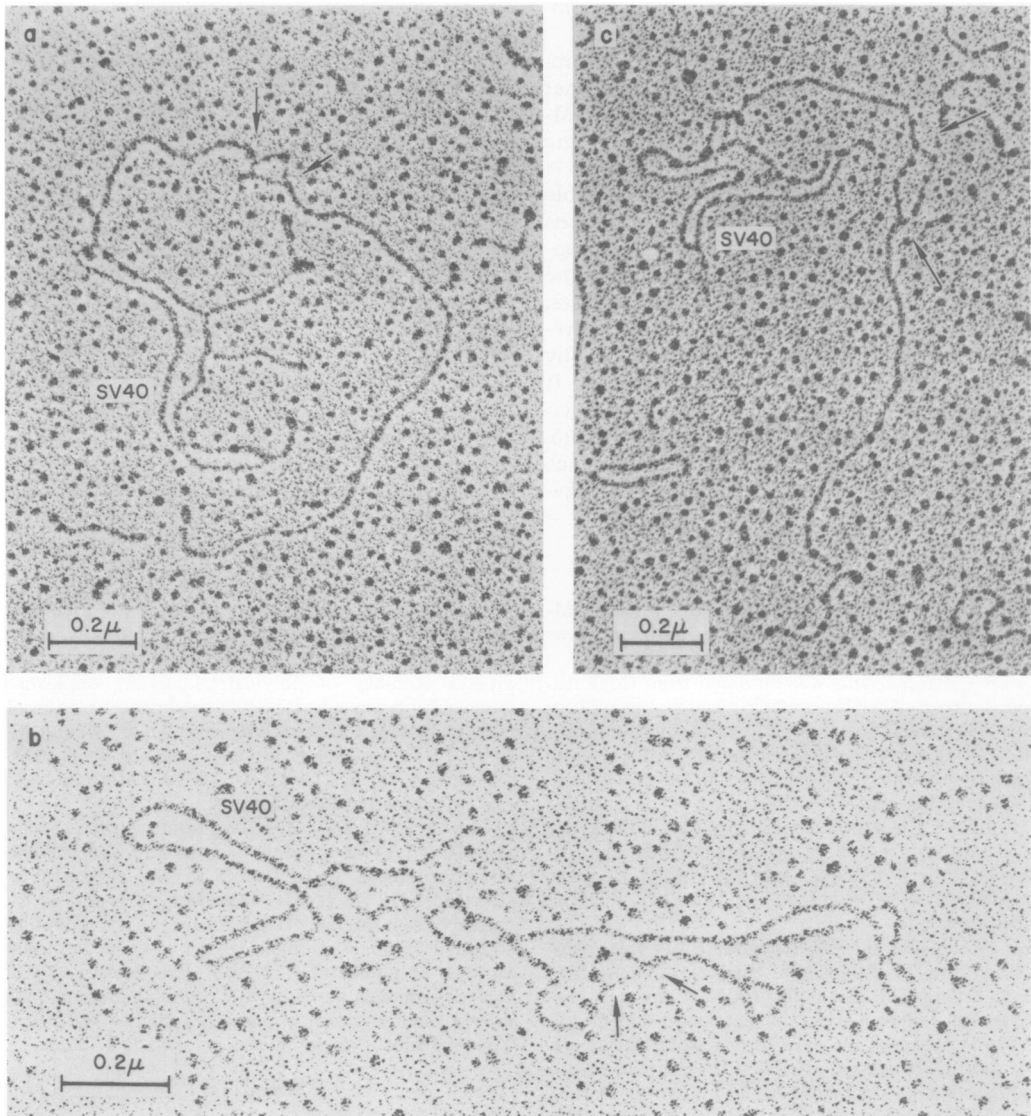


FIG. 1. (a) Linear heteroduplex molecule formed between AKR cDNA and AKR MCF 247 RNA; (b) circular heteroduplex molecule formed between AKR cDNA and AKR MCF 247 RNA; (c) linear heteroduplex molecule formed between M-MLV cDNA and M-MLV₈₃ RNA. A circular SV40-dBrU molecule is attached at the 3' end of the viral RNA for each of the heteroduplex molecules. Two short RNA molecules with intact 3' ends are also attached to the SV40-dBrU molecule in (a). Arrows indicate the beginnings and the ends of the substitution loops (sequence nonhomology regions) discussed in the text.

and AKR MCF 247 RNA is shown in Fig. 1a. The only region of sequence nonhomology between these two viral genomes is a substitution loop. A histogram of the frequency with which a particular point, measured relative to the SV40-dBrU-labeled 3' end, occurs in a nonhomology loop (Fig. 2) shows that the sequence nonhomology region extends from 1.93 ± 0.15 to 2.72 ± 0.16 kb from the 3' end of the RNA. The two arms of the substitution loop appear to be of different lengths; one is measured as $0.7 \pm$

0.16 kb ($n = 48$) and the other as 0.56 ± 0.16 kb ($n = 48$). We believe that this is not a real difference in molecular length but is due to the fact that one arm of the substitution loop is cDNA and the other one is RNA. Under 60% formamide spreading conditions, RNA is more condensed than single-stranded DNA and therefore appears shorter. To extend the RNA as well as the DNA, the hybrids were treated with 1 M glyoxal-0.01 M KPO₄, pH 7.5, at 37°C for 30 min (16) and then spread from 60% formamide. The

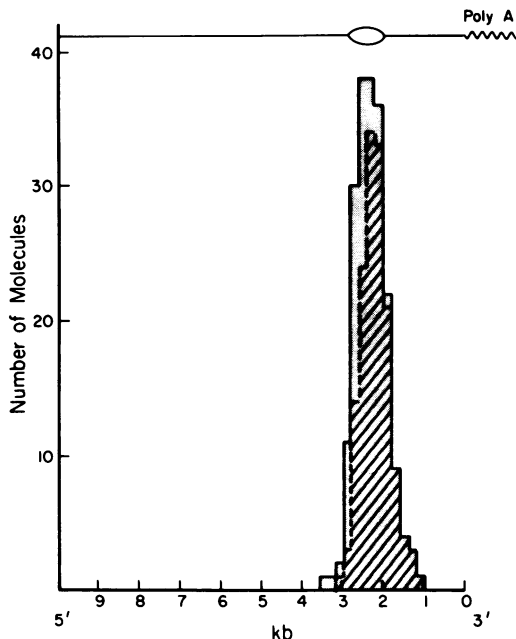


FIG. 2. Histogram of coordinates of the nonhomology loop of AKR/AKR MCF247 heteroduplexes. Heteroduplexes were formed between AKR cDNA and AKR MCF 247 RNA. Sequence homology (double-stranded) regions and sequence nonhomology (single-stranded) regions were measured and converted to units of kilobases (1,000 nucleotides) according to double-strand and single-strand length standards. The frequency of a particular point, measured relative to the SV40-dBrU-labeled 3' end, occurring in a sequence nonhomology loop is plotted against the position along the viral genome. The dotted area is for the longer arm and the hatched area is for the shorter arm of the substitution loop. As discussed in the Results section, the longer arm of the substitution loop is cDNA, and the shorter one is RNA. RNA is more condensed under our spreading conditions and therefore appears shorter. The drawing above the histogram depicts the overall interpretation of the heteroduplex data.

major sequence nonhomology loop now extends from 1.6 to 3 kb from the 3' end of the genomic RNA. The lengths of the longer and shorter arms of the substitution loop were measured as 1.21 ± 0.44 kb ($n = 21$) and 1.37 ± 0.44 kb ($n = 21$), respectively. These numbers are not significantly different. Since glyoxal treatment may result in denaturation of duplex regions bordering the substitution loop (14), we believe that the length (0.70 ± 0.16 kb) attributed to the cDNA strand in the 60% formamide spreads without glyoxal represents the true length of the nonhomology region, and therefore the coordinates of the nonhomology region on both the AKR cDNA and the AKR MCF 247 are 1.93 ± 0.15 to 2.72 ± 0.16 kb.

We did not observe any other nonhomology regions in the AKR cDNA/AKR MCF 247 RNA heteroduplexes. It should be noted that the cDNA molecules varied in length from 6 to 10 kb. Furthermore, not all the RNA molecules were of genome length. Eighty-two percent of the heteroduplexes were of a length greater than 5 kb, 65% of them were of a length greater than 7 kb, and only 12% of them were longer than 9 kb. With the cytochrome *c* spreading technique, single-stranded and double-stranded nucleic acids can not always be unambiguously distinguished. Therefore, even for a heteroduplex molecule with a length greater than 9 kb, part of the molecule can still be single stranded. Since we use 3' end poly(A) selection for the RNA, the amount of heteroduplex data about regions near the 3' end is greater than that for the region near the 5' end of the genome. Therefore, we feel that our data do not completely eliminate the possibility of small nonhomology regions between 6 and 10 kb, especially close to the 5' end of the genome.

Most cDNA synthesis initiates at a tRNA primer close to the 5' end of the RNA template (11), and chain elongation continues from the 3' end of the genome. Therefore, the cDNA transcript is a circular permutation of the viral RNA. As previously discussed (12, 14), this permits the formation of circular heteroduplexes. Such a molecule for the AKR/AKR MCF 247 heteroduplex is shown in Fig. 1b. This result shows that there is sufficient homology between the cDNA 5' sequences of the two RNAs to permit efficient cross-hybridization at the heteroduplex hybridization criterion.

M-MLV/M-MLV₈₃ heteroduplex. A typical heteroduplex molecule formed between M-MLV cDNA and M-MLV₈₃ RNA is shown in Fig. 1c. The histogram tabulating the measurements is shown in Fig. 3. The only sequence nonhomology region is a substitution loop extending from 1.97 ± 0.14 to 2.79 ± 0.14 kb from the 3' end. The two arms are 0.86 ± 0.17 kb ($n = 36$) and 0.64 ± 0.16 kb ($n = 36$). As in the case of AKR/AKR MCF 247, when heteroduplex molecules are treated with glyoxal, the sequence nonhomology region expands to cover the region from 1.7 to 2.8 kb from the 3' end, and appears as a substitution loop with two arms 0.95 ± 0.22 kb ($n = 21$) and 1.04 ± 0.22 kb ($n = 21$). We thus conclude that the sequence nonhomology region between these two viral genomes extends from 2 to 2.8 kb from the 3' end, and the two arms have an approximately equal length of 0.86 kb. In this study, 94, 71, and 37% of the heteroduplex molecules had lengths greater than 5, 7, and 9 kb, respectively. Circular molecules were also observed (data not shown).

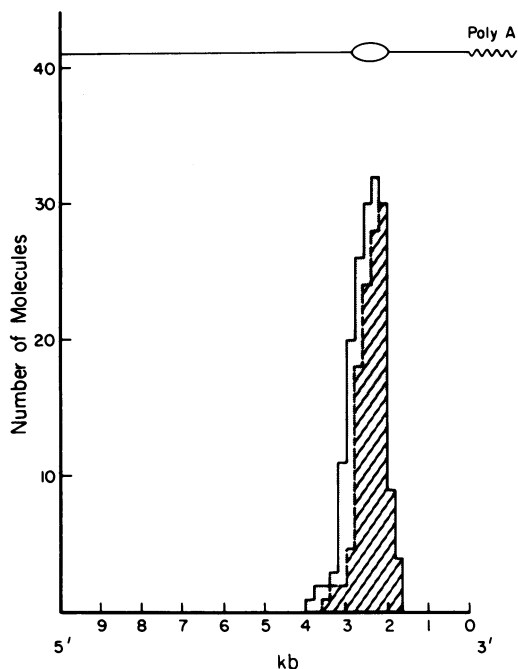


FIG. 3. Histogram of coordinates of the nonhomology loop of *M-MLV/M-MLV₈₃* heteroduplexes. Heteroduplexes were formed between *M-MLV* cDNA and *M-MLV₈₃* RNA. Sequence homology (double-stranded) regions and sequence nonhomology (single-stranded) regions were measured and converted to units of kilobases (1,000 nucleotides) according to double-strand and single-strand size standards. The frequency of a particular point, measured relative to the SV40-dBrU-labeled 3' end, occurring in a sequence nonhomology loop is plotted against the position along the viral genome. The dotted area is for the longer arm and the hatched area is for the shorter arm of the substitution loop. As discussed in the Results section, the longer arm of the substitution loop is cDNA, and the shorter one is RNA. RNA is more condensed under our spreading conditions and therefore appears shorter. The drawing above the histogram depicts the overall interpretation of the heteroduplex data.

M-MLV/AKR heteroduplex. A typical heteroduplex formed between *M-MLV* cDNA and *AKR* RNA is shown in Fig. 4a. A histogram of the results is shown in Fig. 5. There are four substitution loops measured, with coordinates from the labeled 3' end as follows: loop A from 0.89 ± 0.10 to 1.25 ± 0.16 kb, with two arms with lengths of 0.47 ± 0.17 kb ($n = 94$) and 0.33 ± 0.14 kb ($n = 94$); loop B from 2.01 ± 0.13 to 2.55 ± 0.17 kb, with two arms with lengths of 0.56 ± 0.24 kb ($n = 90$) and 0.36 ± 0.15 kb ($n = 90$); loop C from 3.03 ± 0.17 to 3.45 ± 0.22 kb, with two arms with lengths of 0.48 ± 0.18 kb ($n = 84$) and 0.32 ± 0.13 kb ($n = 84$); loop D from 6.55 ± 0.20 to 7.07 ± 0.41 kb, with two arms with

lengths of 0.43 ± 0.17 kb ($n = 29$) and 0.36 ± 0.15 kb ($n = 29$). The pattern of the three loops on the 3' half of the molecule was rather reproducible. However, the positions of the loops and the spacers between them were somewhat variable and there were some molecules in which loop B and/or C was not present. Furthermore, when spread from a more denaturing solvent, for example, 80% urea + formamide instead of 60% urea + formamide, more than 90% of the molecules observed had the three loops. When spread from a less denaturing solvent, 50% urea + formamide instead of 60% urea + formamide, 75% of the molecules observed had only loop A. These results indicate that there is partial sequence mismatch between the two genomes in the region from 0.9 to 3.5 kb from the 3' end.

From the histogram, the frequency of occurrence of loop D is lower than that of loops A, B, and C, and its measured position is more variable; nevertheless, we think it is real for the following reasons. (i) Not every hybrid molecule is full length (103 molecules have lengths greater than 5 kb; only 40 molecules have lengths greater than 8 kb). However, the percentage (about 35%) of molecules with lengths greater than 8 kb that have loop D is comparable to the percentage (about 44%) of molecules with lengths greater than 5 kb that have loops A, B, and C. (ii) We determined the position of loop D by summing the length of a series of double-stranded and single-stranded regions measured from the 3' end. Since the uncertainties in length measurements accumulate, the position of loop D cannot be determined as accurately as those of loops A, B, and C, and appears to be more variable.

As in the previously described cases, circular heteroduplex molecules also were observed. An example is shown in Fig. 4b.

When *M-MLV* cDNA was hybridized to 35S *M-MLV* RNA, and when *AKR* cDNA was hybridized to *AKR* RNA, no molecules resembling those shown in Fig. 1a, 1b, 1c, 4a, and 4b were observed. When cDNA's or RNAs were incubated separately under the conditions used for heteroduplex formation, no heteroduplex structures were found. These control experiments eliminate the possibility that the structures observed are artifacts caused by cellular DNA that may have copurified with the viral cDNA, or that the parental RNAs are themselves heterogeneous in sequence.

DISCUSSION

The MCF strains of MLVs appear to play a key role in the genesis of spontaneous lymphomas (10). These viruses manifest a dual host range and appear to be recombinants between

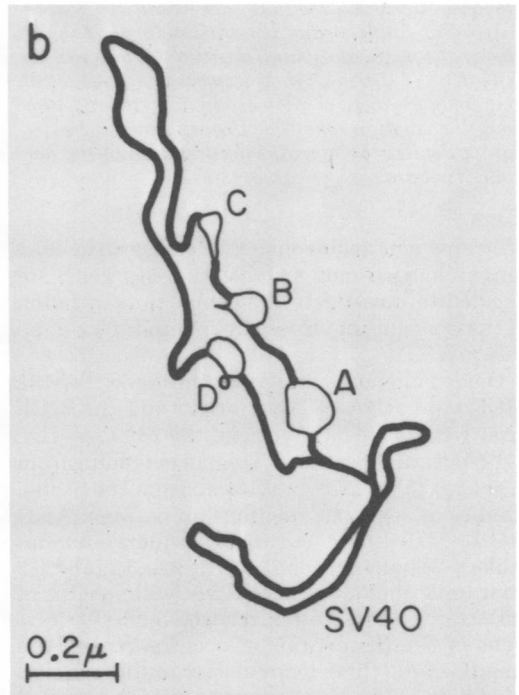
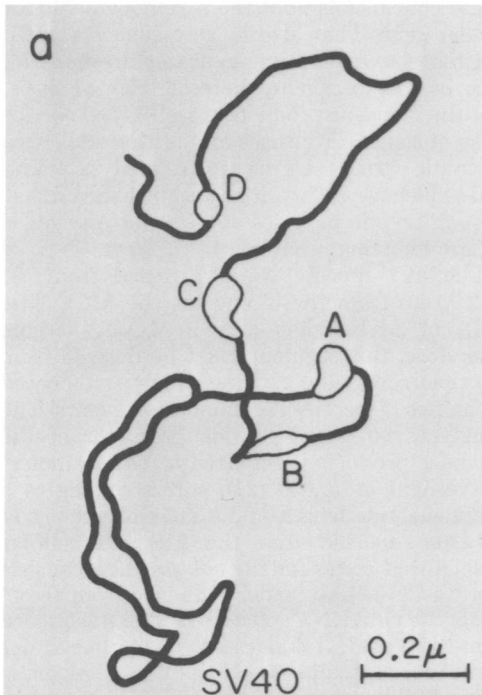
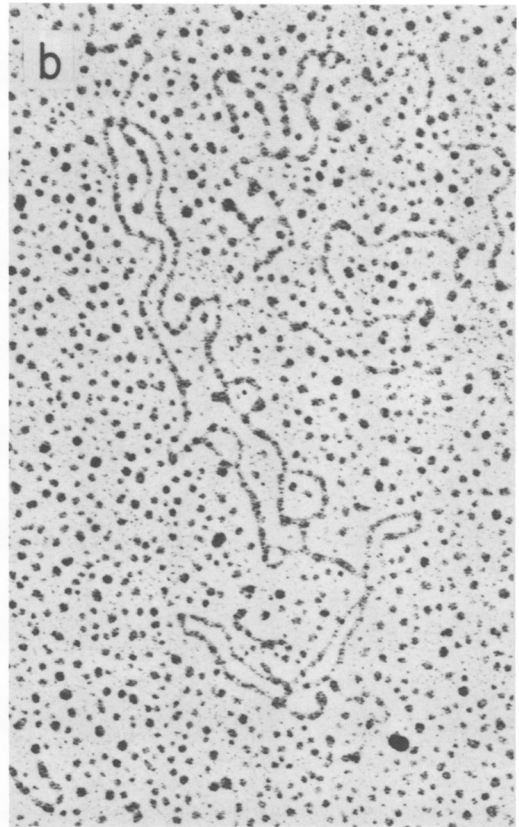
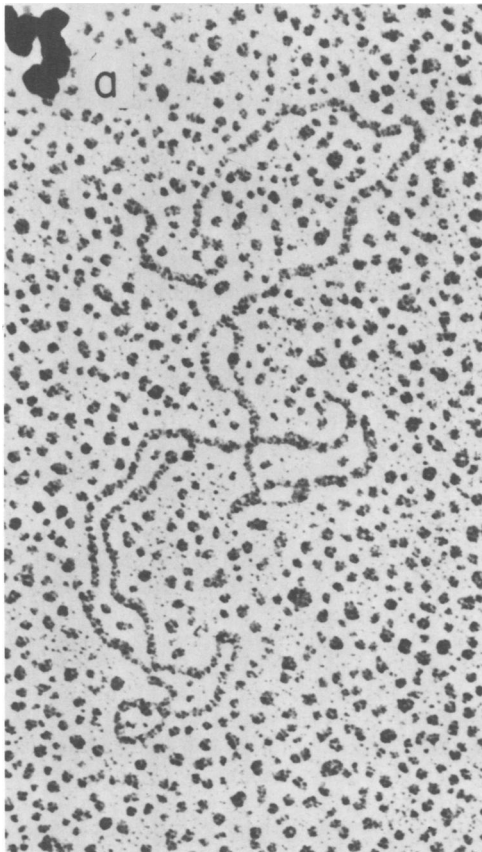


FIG. 4. Electron micrographs of M-MLV/AKR heteroduplexes. (a) Linear heteroduplex molecule formed between M-MLV cDNA and AKR RNA. (b) Circular heteroduplex molecule formed between M-MLV cDNA and AKR RNA. Circular SV40-dBrU molecules are attached at the 3' end of the viral RNA in the heteroduplexes. Interpretative tracings of the heteroduplexes are presented below the micrographs. Heavy lines indicate duplex regions; thin lines represent single-stranded regions.

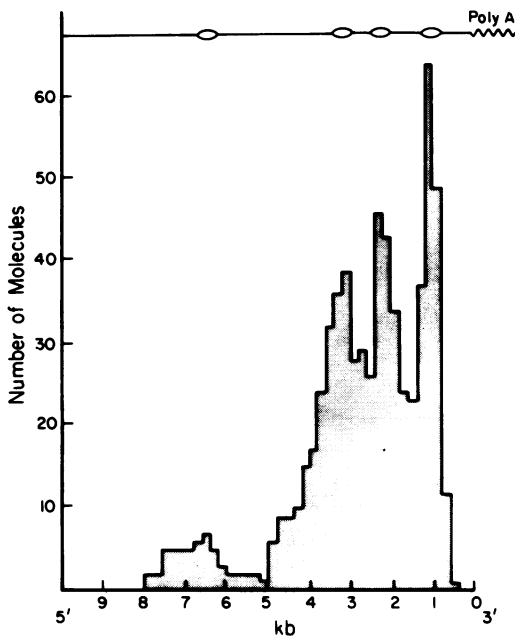


FIG. 5. Histogram of coordinates of the nonhomology loops of M-MLV/AKR heteroduplexes. Heteroduplexes were formed between M-MLV cDNA and AKR RNA. Sequence homology (double-stranded) regions and sequence nonhomology (single-stranded) regions were measured and converted to units of kilobases (1,000 nucleotides) according to double-strand and single-strand length standards. The frequency of a particular point, measured relative to the SV40-dBrU-labeled 3' end, occurring in a sequence nonhomology loop is plotted against the position along the viral genome. The drawing above the histogram depicts the overall interpretation of the heteroduplex data.

ecotropic and xenotropic MLVs. Since the host range changes may involve the *env* gene, we decided to investigate the genome organization of these strains of viruses by heteroduplex analysis.

Our results are summarized in Fig. 6. For the MCF viral RNA of AKR origin and the MCF viral RNA of M-MLV origin, the MCF-specific sequences are located in a region extending from 1.95 (± 0.15) to 2.75 (± 0.15) kb from the 3' end. Studies of sequence relationship between AKR and M-MLV show four major sequence nonhomology regions centered at 1.1, 2.3, 3.3, and 6.8 kb from 3' end of the viral RNA, with lengths of 0.5, 0.6, 0.5, and 0.5 kb, respectively. Observations of the frequencies of occurrence and the lengths of the three loops closest to the 3' end as a function of the denaturing power of the spreading solution and the hypophase suggest that there is some degree of partial sequence mismatch in the entire region from 0.9 to 3.5 kb,

with the greatest extent of nonhomology within loop A, less in loops B and C, and some mismatch in the regions represented as duplex in Fig. 6.

We wish to correlate the results of our physical study with the available information as to the genetic maps of murine viral RNAs.

In vitro translation studies (9, 15, 18), immunoprecipitation of virus-specific RNA from polyosomes by monospecific p30 antibodies (17), and the analysis of the MSV clone 124/M-MLV heteroduplex (12) suggest that the genes coding for the internal structure proteins (*gag*) lie at the 5' end of the viral RNA. The precursor for the *gag* gene products is a 65,000-dalton protein (22). The minimum coding sequence is therefore about 1.8 kb. Furthermore, the precursor for *gag* plus *pol* gene products is a polypeptide with a molecular weight of 180,000 to 200,000, for which translation starts close to the 5' end. The length of the coding sequence is therefore 5 to 5.5 kb.

Rothenberg et al. (20) reported that 0.5 ± 0.2 kb of sequence at the extreme 5' end of the viral RNA is a leader sequence "spliced" to the 21 to 24S *env* mRNA. It is not known whether or not this leader is coding for *gag* proteins. All these considerations suggest that the coding sequences for *gag* extend from 7.2 ± 0.5 to 9.7 ± 0.5 kb from the 3' end and that the *gag* plus *pol* sequences extend from 3.7 ± 0.5 to 9.7 ± 0.5 kb. These conclusions limit the region available for the *env* gene. They also suggest, since the MCF and respective ecotropic genomes are homologous in this region by heteroduplex analysis, that the *gag* gene products as well as the polymerase of the MCF viruses and of the respective ecotropic viruses are closely related in amino acid sequences. In fact, it has been observed that the *gag* tryptic peptides of Akv and Akv MCF 247 are indistinguishable (4).

The MCF-specific sequence extends from 1.95 to 2.75 kb from the 3' end for the MCF viral RNA of AKR origin and of M-MLV origin. Therefore, this segment must be derived from the xenotropic viral genome which participated in the two respective recombination events. The nonglycosylated polypeptide precursor of the *env* gene products is reported to have a molecular weight of 70,000 (24), corresponding to a polynucleotide length of 1.9 kb. Rothenberg et al. (20) reported that the 21S intracellular mRNA that codes for the *env* proteins consists of a 0.5 (± 0.2)-kb leader sequence from the 5' end of the viral RNA spliced to a 2.8-kb sequence from the 3' end. It is not known whether or not the leader sequence is coding. The 3' terminus of the 1.9-kb precursor *env* gene must therefore lie at 1.4 kb or less from the 3' end. The heteroduplex results therefore predict that the carboxyl-terminal amino acid sequences of the

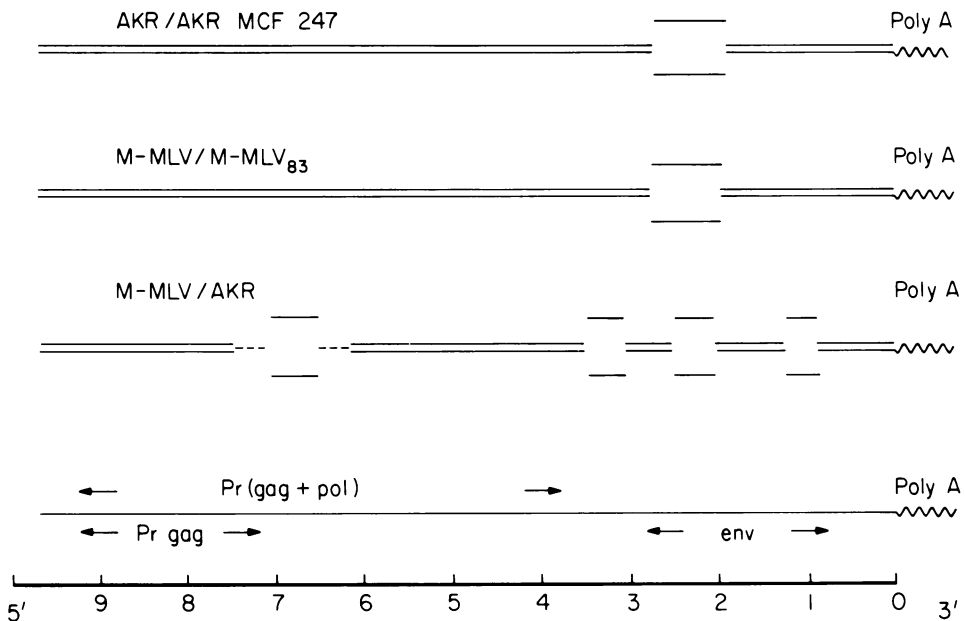


FIG. 6. Summary of heteroduplex structures for AKR/AKR MCF 247, M-MLV/M-MLV₈₃, and M-MLV/AKR. Heteroduplexes were formed between AKR cDNA and AKR MCF 247 RNA, between M-MLV cDNA and M-MLV₈₃ RNA, and between M-MLV cDNA and AKR RNA. Double lines indicate regions of sequence homology; single lines above or below the duplex double lines indicate sequences present in one virus and not in the other (sequence nonhomology regions). The dotted lines in the left-most nonhomology region of M-MLV/AKR heteroduplex indicate the variability in the measured position of that loop, as discussed in the text.

MCF *env* precursor polypeptides are similar (or identical) to those of the ecotropic parents, whereas the MCF-specific sequences from 1.95 to 2.75 kb must contain the xenotropic polypeptide part of the mature gp70.

The glycoproteins of AKR and of M-MLV differ in their tryptic peptide maps (5) and are not closely related by immunological criteria (13). The major sequence nonhomology regions of these two viral genomes were found between 0.9 and 3.5 kb from the 3' end. These results further support the notion that the *env* gene is located within this region and make it unlikely that the 0- to 1-kb region, where AKR and M-MLV are relatively homologous, code for any significant length of gp70.

The sequence nonhomology region D (Fig. 6) between AKR and M-MLV with coordinates of 6.55 ± 0.20 to 7.07 ± 0.41 kb appears to be within the gene coding for the *gag-pol* precursor protein. Whether it lies in the *gag* gene or *pol* gene is not known.

It is rather striking that the coordinates of the two junctions of MLVs with MCF-specific sequences are approximately the same for both the AKR/AKR MCF and the M-MLV/M-MLV₈₃ cases. It should be noted that the heteroduplex study shows that the AKR and M-MLV sequences adjacent to the 1.95- and 2.75-kb junc-

tions with MCF specific sequences are relatively homologous (Fig. 6). It may be that these similar sequences are hot spots for recombination in each case.

Rommelaere et al. (19) have constructed maps of RNase T1-resistant oligonucleotides obtained from AKR MCF 247, and Shih et al. (21) have mapped large RNase T1-resistant oligonucleotides of M-MLV₈₃. Our results obtained by heteroduplex analysis are in general agreement with those obtained by oligonucleotide fingerprinting techniques. Their findings indicate that the major sequence nonhomology regions between an MCF strain of virus and its parental MLV genome are clustered at the 3' half of the RNA. Fingerprinting analysis of MCF AKR 247 RNA showed an additional oligonucleotide (MCF 247 oligonucleotide 107) about 14 nucleotides long, specific for MCF virus which is located in the 5' half of the genome. Similarly, analysis of M-MLV₈₃ RNA also showed a specific spot of about 23 nucleotides (oligonucleotide A5) identified in the 5' half of the genome. The heteroduplexes shown here, however, do not detect any sequence nonhomology in this region. This is probably due to the fact that, under our conditions for spreading, 10 to 20% random sequence mismatch will show as a sequence homology region. We also failed to detect a sequence nonhomology

feature corresponding to the AKR MCF 247 specific oligonucleotide (19 nucleotides long) at the very 3' end of the viral RNA (19). However, under the conditions of the cytochrome *c* spreading technique, a nonhomology feature with a length under 50 nucleotides cannot be detected unambiguously.

Thus, the physical evidence from heteroduplex structures is consistent with the hypothesis, derived from biological properties and tryptic peptide analysis, that the MCF viruses are recombinant within the *env* gene.

Since completing this work and manuscript, we have learned that D. J. Donoghue and collaborators at the Massachusetts Institute of Technology have observed a very similar heteroduplex structure between M-MLV and the dual host range HIX virus as observed here for the MLV/MCF viruses (personal communication).

ACKNOWLEDGMENTS

We thank Fred Jensen of Scripps Clinic and Research Foundation for a gift of AKR MCF 247 virus and for advice and assistance, and Billie Green and Marianne McKenett for technical assistance.

Y.-H.C. has been supported in part by postdoctoral fellowship F32CA05099 from the National Institutes of Health. This research was supported by contract NO1 CP 43306 with the Virus Cancer Program of the National Cancer Institute (N.D.) and by Public Health Service grants CA 16561 and 21408 from the National Cancer Institute.

LITERATURE CITED

- Bailey, J. M., and N. Davidson. 1976. Methylmercury as a reversible denaturing agent for agarose gel electrophoresis. *Anal. Biochem.* **70**:75-85.
- Bender, W., and N. Davidson. 1976. Mapping of poly(A) sequences in the electron microscope reveals unusual structure of type C oncornavirus RNA molecules. *Cell* **7**:595-607.
- Davis, R., M. Simon, and N. Davidson. 1971. Electron microscopic heteroduplex methods for mapping regions of base sequence homology. *Methods Enzymol.* **21**:413-428.
- 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.
- Elder, J. H., F. C. Jensen, M. L. Bryand, 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.
- Fan, H., and D. Baltimore. 1973. RNA metabolism of murine leukemia virus: detection of virus-specific RNA sequences in infected and uninfected cells and identification of virus-specific messenger RNA. *J. Mol. Biol.* **80**:93-117.
- Fan, H., and M. Paskind. 1974. Measurement of the sequence complexity of cloned Moloney murine leukemia virus 60 to 70S RNA: evidence for haploid genome. *J. Virol.* **14**:421-429.
- Fischinger, P. J., S. Nomura, and D. P. Bolognese. 1975. A novel murine oncornavirus with dual eco- and xenotropic properties. *Proc. Natl. Acad. Sci. U.S.A.* **72**:5150-5155.
- Gielkens, A. L. J., D. Van Zaane, H. P. J. Bloemers, and H. Bloemendal. 1976. Synthesis of Rauscher murine leukemia virus-specific polypeptide *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* **73**:356-360.
- 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.
- Haseltine, W. A., D. G. Kleid, A. Panet, E. Rothenberg, and D. Baltimore. 1976. Ordered transcriptions of RNA tumor virus genomes. *J. Mol. Biol.* **106**:109-131.
- Hu, S., N. Davidson, and I. M. Verma. 1977. A heteroduplex study of the sequence relationships between the RNAs of M-MSV and M-MLV. *Cell* **10**:469-477.
- Ihle, J. N., T. P. Penny, and D. P. Bolognese. 1976. Purification and serological characterization of the major envelope glycoprotein from AKR murine leukemia virus and its reactivity with autogenous immune sera from mice. *J. Virol.* **17**:727-736.
- Junghans, R. P., S. Hu, C. A. Knight, and N. Davidson. 1977. Heteroduplex analysis of avian RNA tumor viruses. *Proc. Natl. Acad. Sci. U.S.A.* **74**:477-481.
- Kerr, L. M., U. Olshevsky, H. F. Lodish, and D. Baltimore. 1976. Translation of murine leukemia virus RNA in cell-free systems from animal cells. *J. Virol.* **18**:627-635.
- Kung, H. J., S. Hu, W. Bender, J. M. Bailey, N. Davidson, M. O. Nicolson, and R. M. McAllister. 1976. RD114, baboon, and woolly monkey viral RNAs compared in size and structure. *Cell* **7**:609-620.
- Mueller-Lantzsch, N., and H. Fan. 1976. Monospecific immunoprecipitation of murine leukemia virus polyribosomes: identification of p30 protein-specific messenger RNA. *Cell* **9**:579-588.
- Philpson, L., P. Anderson, 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.
- Rommelaere, J., D. V. Faller, and N. Hopkins. 1978. Characterization and mapping of RNase T1-resistant oligonucleotides derived from the genomes of Akv and MCF murine leukemia viruses. *Proc. Natl. Acad. Sci. U.S.A.* **75**:495-499.
- Rothenberg, E., D. J. Donoghue, and D. Baltimore. 1978. Analysis of a 5' leader sequence on murine leukemia virus 21S RNA. Heteroduplex mapping with long reverse transcriptase products. *Cell* **13**:435-451.
- Shih, T. Y., M. O. Weeks, D. H. Troxler, J. M. Coffin, and E. M. Scolnick. 1978. Mapping host range-specific oligonucleotides within genomes of the ecotropic and mink cell focus-inducing strains of Moloney murine leukemia virus. *J. Virol.* **26**:71-83.
- Stephenson, J. R., S. R. Tronick, and S. A. Aaronson. 1975. Murine leukemia virus mutants with temperature-sensitive defects in precursor polypeptide cleavage. *Cell* **6**:543-548.
- Troxler, D. H., D. Lowy, R. Howk, H. Young, and E. M. Scolnick. 1977. The formation of an acute leukemia virus, the spleen focus-forming virus, was accompanied by recombination between ecotropic murine type-C virus and the *env* gene region of xenotropic type-c virus. *Proc. Natl. Acad. Sci. U.S.A.* **74**:4671-4675.
- Van de Ven, W. J. M., C. Onnekink, A. J. M. Vermorcken, and H. P. J. Bloemers. 1977. Effect of impaired glycosylation on the synthesis of envelope proteins of Rauscher murine leukemia virus. *Virology* **8**:334-344.
- Verma, I. M. 1978. Genome organization of RNA tumor viruses I. *In vitro* synthesis of full-genome-length single-stranded and double-stranded viral DNA transcripts. *Virology* **26**:615-629.