JOURNAL OF VIROLOGY, Dec. 1985, p. 691–700
0022-538X/85/120691-10$02.00/0

Vol. 56, No. 3

Normal Expression of Polymorphic Endogenous Retroviral RNA Containing Segments Identical to Mink Cell Focus-Forming Virus†

DAVID E. LEVY,[1,2]‡ RICHARD A. LERNER,[2] AND MICHAEL C. WILSON[2]*

Division of Biology, California Institute of Technology, Pasadena, California 91125,[1] and Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037[2]

Received 24 April 1985/Accepted 16 August 1985

In the absence of infectious virus, strains of mice express polyadenylated RNA transcripts homologous to the genome of murine leukemia virus. In addition to transcripts consistent with full-length and spliced *env* retroviral RNAs, several unique RNA species which lack the *env* sequence accumulate in a tissue-specific manner. These RNA species are presumed to be transcribed from endogenous retroviral sequences that constitute the bulk of the murine leukemia virus-related sequences in the murine genome. To determine the relationship of these RNA transcripts to infectious murine leukemia virus and the precise structural basis of the heterogeneity observed for the *env*-lacking transcripts, we isolated and sequenced cDNA recombinants representing the RNAs expressed in strain 129 G_{IX}^+ mice. Comparisons of the nucleotide sequences demonstrated that the endogenous retroviral transcripts differed in *pol*, p15E, and R-peptide regions by single nucleotide changes. In contrast, the gp70-coding regions of two cDNA clones derived from epididymis and liver were completely homologous over a 599-nucleotide overlapping sequence. The structures of *env*-lacking transcripts were examined in two independent cDNA clones, and each was found to contain a different deletion that was potentially mediated by seven-base pair direct repeats in the intact sequence. The extensive sequence homology between cDNAs allowed construction of a cumulative sequence map of the 3′ end of an intact endogenous retroviral transcript. A comparison of this sequence with infectious ecotropic and mink cell focus-forming viruses revealed that the endogenous transcripts are highly homologous with the substituted portions of leukemogenic mink cell focus-forming viruses and therefore further define the boundaries of recombination required to generate these viruses.

Infectious type C retroviruses of inbred mice display three distinct host ranges, which are mediated through their viral envelope glycoproteins by cell receptor recognition. Ecotropic, xenotropic, and dualtropic viral isolates have been classified on the basis of the ability to infect primarily mouse cells, only nonmouse cells, and cells of both mouse and nonmouse origin, respectively. Members of the third class, the dualtropic viruses, are more commonly referred to as mink cell focus-forming (MCF) viruses because of their ability to cause characteristic alterations in the cellular morphology of infected mink cells. Infectious ecotropic and xenotropic viruses can be induced directly from proviruses integrated within the genomes of various mouse strains. However, MCF viruses are formed de novo by recombination between an exogenous, infectious ecotropic virus and endogenous viral sequences related to xenotropic virus, resulting in the replacement of variable portions of the 3′ end of the ecotropic genome with endogenous information encoding altered envelope proteins. It is viruses of this type which are generally associated with the development of leukemia in mice (see reference 38 for a review).

Endogenous viral sequences, distinct from ecotropic and xenotropic proviruses, are present in multiple copies in the murine genome (7, 17, 19); however, their transcriptional activity in normal, nonviremic mice has not been fully documented. Even in the absence of the ability to produce infectious virus, strains of mice express antigens thought to be encoded from such endogenous viral sequences (reviewed in reference 26). The precise relationship between actively transcribed viral sequences and sequences involved in the generation of recombinant MCF viruses is not known. In different strains considerable heterogeneity exists in the levels and structures of the viral proteins, and there is evidence that this expression is regulated in specific tissues (32). This regulation of endogenous retroviral gene expression appears to be distinct from the genetically defined loci governing the inducibility of infectious virus, which include both proviral structural genes themselves and regulatory genes distinct from the induced virus (15, 20, 21).

Genetic analysis of endogenous retroviral gene expression has been greatly facilitated by the generation of congenic partner strains of strain 129 mice, G_{IX}^+ and G_{IX}^-, which are characterized by high and low levels of virus-related antigen production, respectively (43, 44). Recently, we reported that expression of multiple but distinct tissue-specific RNA transcripts which are thought to encode these endogenous viral antigens is under coordinate regulation by a *trans*-acting product of the *Gv-1* locus (25).

To determine the structural basis for this heterogeneity, to document the number of independent transcription units that encode these transcripts, and to define the relationship between the normal cellular components and the products of exogenous viruses, the nucleotide sequences of cDNA clones corresponding to endogenous virus-related transcripts expressed in strain 129 G_{IX}^+ mice were determined. The polymorphism previously detected in the RNA population was found to be due to deletions of varying lengths of retroviral sequences accompanied by scattered single nucleotide differences between different transcripts. These findings demonstrate the activity of multiple genomic transcription units in the origin of these regulated RNA species.

* Corresponding author.
† Publication 3655-MB of the Department of Molecular Biology, Scripps Clinic and Research Foundation.
‡ Present address: Rockefeller University, New York, NY 10021.

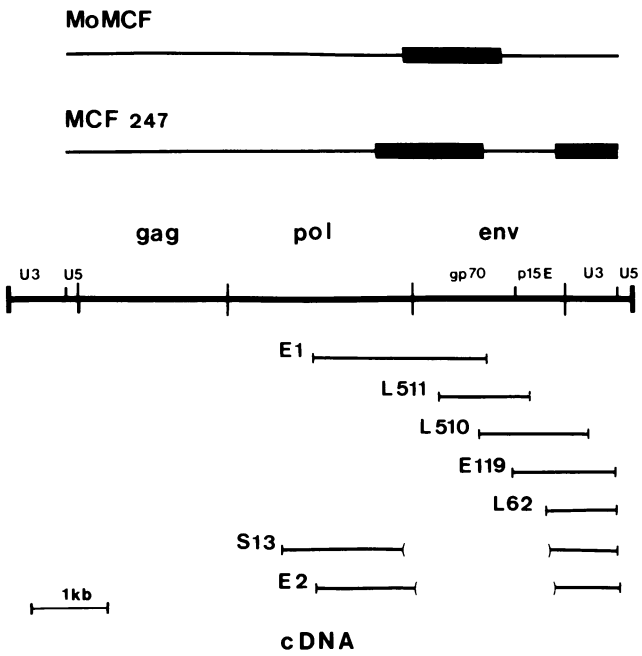

FIG. 1. Structure of retroviral cDNA clones. The cDNA insertions of retroviral clones derived from strain 129 G_{IX}^+ polyadenylated RNA were aligned with the Mo-MuLV proviral structure (middle) by restriction enzyme mapping, hybridization, and DNA sequencing. The vertical bars in the cDNA clones indicate the positions of junctions of eucaryotic sequences with plasmid DNA. Parentheses enclose deleted regions of the retroviral sequence. The designation of each clone includes an initial letter which indicates the tissue from which the RNA was derived, as follows: E, epididymis; L, liver; S, spleen. At the top are diagrams of Mo-MCF and MCF 247 viral genomic RNAs. The thin lines represent regions of ecotropic sequences, while the thick lines represent regions of nonecotropic information substituted into the genome by recombination. The MCF 247 virus diagram is based on the restriction maps of Chattopadhyay et al. (7) and the sequences described by Kelley et al. (16) and Holland et al. (14). The diagram of Mo-MCF is based on the data of Bosselman et al. (6). kb, Kilobase.

Furthermore, the structure of these transcripts establishes that the normal cellular components are most likely encoded by endogenous proviruses rather than by homologous cellular genes, as opposed to the cellular counterparts of the oncogenes acquired by acutely transforming retroviruses (46). Comparisons with known viral sequences have revealed a striking similarity between the viral RNA transcripts normally expressed in strain 129 mice and the sequences of the substituted portions of leukemogenic dualtropic MCF viruses. These similarities delineate the minimal recombination points necessary to create MCF virus and suggest a possible relationship between transcriptionally active proviruses and the generation of recombinant viruses. Protein products translated from the *env* sequence reported here should display the dualtropic determinants associated with MCF virus gp70.

(This research was conducted by D. E. Levy in partial fulfillment of the requirements for a Ph.D. degree from California Institute of Technology, Pasadena.)

MATERIALS AND METHODS

The construction and characterization of cDNA libraries from various tissues of strain 129 G_{IX}^+ mice have been described previously (25). Briefly, total cell RNAs were obtained from livers, thymuses, spleens, and epididymides of 6- to 8-week-old male mice by pulverization of frozen tissues, treatment with sodium dodecyl sulfate and proteinase K, and phenol-chloroform extraction. cDNA synthesis was initiated on poly(A)-enriched RNA by oligo(dT) priming, and cDNA was rendered double stranded by treatment with the large fragment of *Escherichia coli* DNA polymerase I and ultimately inserted into pBR322 by G-C tailing. Clones representing virus-related RNA were selected by hybridization with radioactive probes derived from cloned Moloney murine leukemia virus (Mo-MuLV), as described previously (24).

Selected clones were sequenced by the chain termination method of Sanger et al. (40) after subcloning into M13 vectors (30). The resulting sequences were analyzed with the aid of computer programs supplied by the National Biomedical Research Foundation, Washington, D.C. (34).

RESULTS

The transcripts of expressed proviruses in strain 129 mice constitute a polymorphic family of sequences that differ by deletions of viral sequences of variable length (24, 25) accompanied by scattered single base differences among individual members (see below). Although no single cDNA containing a sequence homologous to an entire *pol* or *env* gene open reading frame was isolated (Fig. 1), enough overlapping sequence with a high degree of nucleotide homology was obtained to construct a cumulative sequence map of the 3′ end of an intact viral transcript. This reconstructed sequence allowed comparisons with AKR endogenous proviral clone A-12 (17), AKR MCF 247 virus (14, 16), Moloney MCF (Mo-MCF) virus (6), and AKV ecotropic virus (12, 22).

***pol* gene.** An analysis of restriction enzyme sites and hybridization to cloned Mo-MuLV sequences (3, 24, 25) indicated that clones E1, E2, and S13 contained *pol* region sequences (Fig. 1). The 360-nucleotide sequence from the 3′ end of the *pol* gene of clone E1 and the sequences from the corresponding regions of clones E2 and S13 are shown in Fig. 2 along with the predicted amino acid sequence of clone E1. The sequence heterogeneity of the endogenous retroviral transcripts of strain 129 G_{IX}^+ mice was clearly illustrated by these clones. First, sequence heterogeneity occurred within the overlapping regions of these clones. Of the 315 nucleotides aligned between clones E1 and E2, there were 15 mismatches and one two-nucleotide gap at positions 57 and 58 (a 4.8% difference). Between clones E1 and S13, there were five differences, resulting in differences of 3.2 and 5.3%, respectively. Second, clones E2 and S13 exhibited large deletions which eliminated the 3′ portion of the *pol* gene described here. These deletions extended through the majority of the *env* gene, resulting in fusion of the *pol* sequence with the 3′ end of the p15E coding sequence (Fig. 3). The actual endpoints for these deletions were ambiguous due to seven-base pair direct repeats at these deletion points in the undeleted sequence (25). The deletion in S13 extended from the sequence GGACCCT at positions 178 to 184 to the second occurrence of this sequence at positions 478 to 484 in p15E (Fig. 3), retaining a single copy of the repeat and thereby deleting 1,977 nucleotides. A similar deletion in E2 began at nucleotide 316 (Fig. 2) after the sequence GGTC CAG and likewise extended through the second copy of this repeat at positions 501 to 507 of p15E, thus eliminating 1,817 nucleotides. Transcripts displaying analogous deletions have been detected by S1 mapping in a variety of tissues of strain

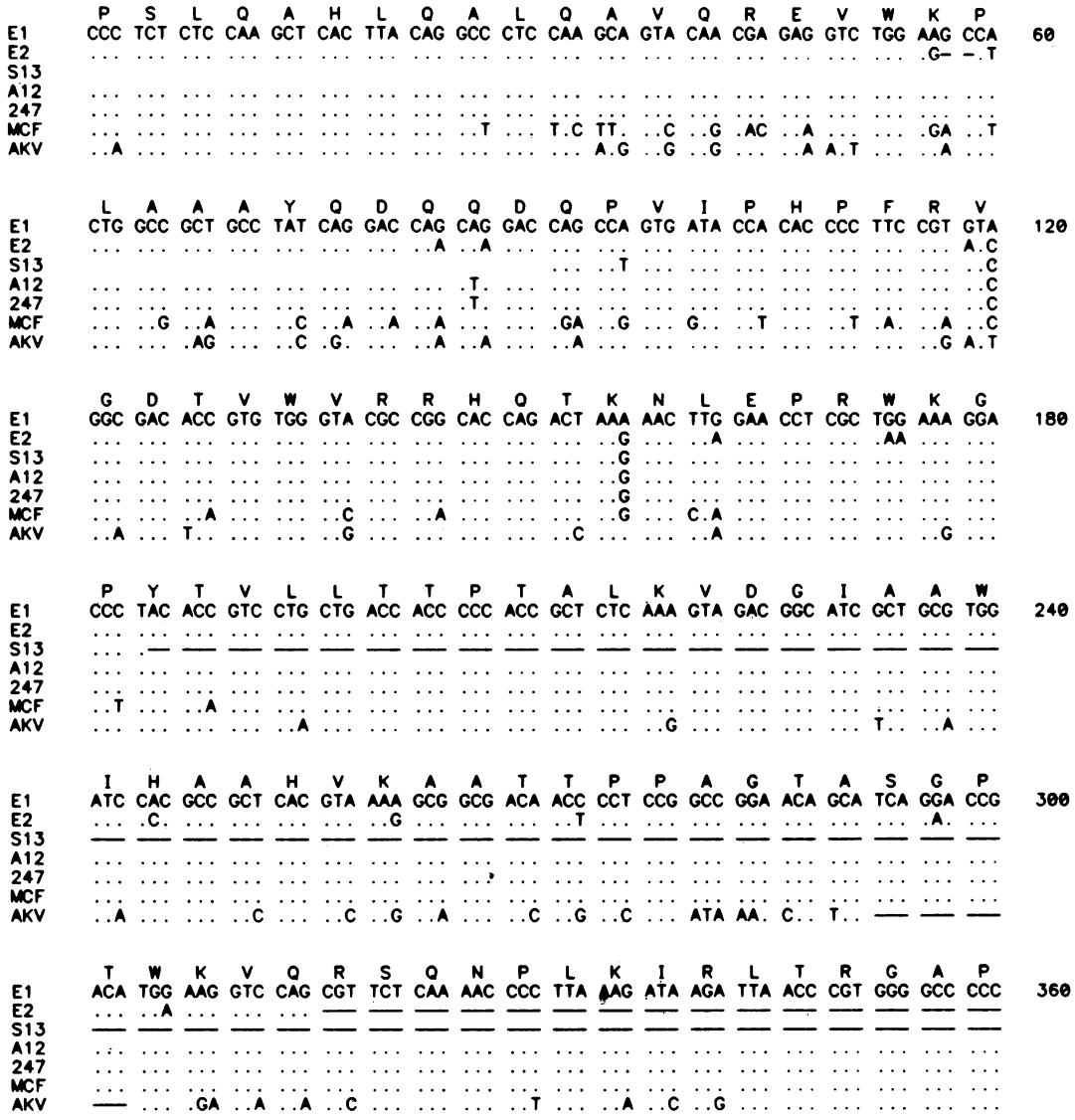

FIG. 2. Sequence comparison of the *pol* region. The DNA sequences and predicted amino acid sequence for 360 nucleotides from the 3′ end of the *pol* gene open reading frame of endogenous transcript clone E1 are shown. This nucleotide sequence is compared with sequences of clones E2 and S13 from strain 129 G_{IX}^+ mice, proviral genomic clone A12 from AKR mice (17), AKR MCF 247 virus (247) from an AKR thymoma (14), Mo-MCF virus (MCF) from a BALB/Mo thymoma (6), and AKV virus, the endogenous ecotropic virus of AKR mice (12, 22). Dots indicate nucleotides identical to nucleotides in the sequence of clone E1, dashes indicate gaps introduced in the sequences to increase the alignment with E1, and blank spaces are present where no sequence data are available. The numbers on the right indicate the nucleotides of the E1 sequence. A single-letter code is used for the translated amino acid sequence, which is indicated above the codons.

129 G_{IX}^+ mice (25). The nucleotide sequence of clone E1 predicted a 120-amino acid open reading frame ending with an ochre codon immediately after nucleotide 360 (nucleotides 59 to 61 [Fig. 4]). This open reading frame overlapped the *env* open reading frame for 58 nucleotides, encoding 19 amino acids of *pol* and 19 amino acids of gp70 in different triplet reading frames (Fig. 4). The methionine initiation codon for *env* began with nucleotide 303 (Fig. 2). Such sequence overlap at the *pol-env* junction is characteristic of murine type C viruses. The sequences of the endogenously expressed RNA also differed with respect to the viral sequences shown in Fig. 2 (Table 1). Clone E1 differed from AKR sequences A12 and MCF 247 (which were identical in this region) in 3 of 360 nucleotides, leading to a single predicted amino acid change (Glu to Leu at positions 85 to 87). E1 differed from Mo-MCF at 36 of 360 nucleotides, resulting in seven predicted amino acid differences, and varied from AKV at 53 of 348 nucleotides, requiring introduction of a gap of four codons to increase sequence similarity. It should be noted that Mo-MCF contained both ecotropic and endogenous information in this region of *pol* since the recombination probably took place at position 210 (Fig. 2) (6). All of the differences between E1 and Mo-MCF occurred 5′ to this probable recombination point between Mo-MCF and the parental virus, Mo-MuLV. Thus, the substituted regions of *pol* in both of these recombinant MCF viruses were virtually identical to the E1 sequence, even though these viruses arose in different mouse strains (AKR

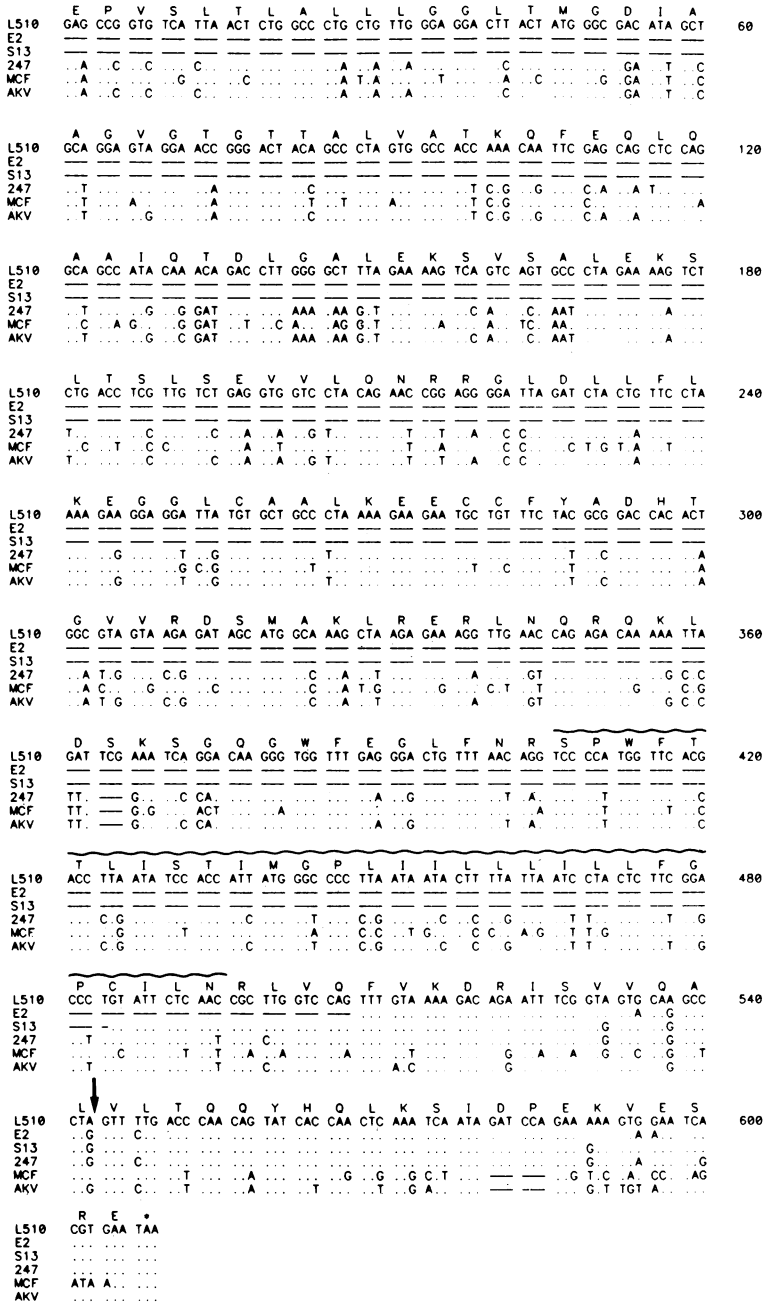

FIG. 3. Sequence comparison of the p15E-coding region. The DNA sequence and predicted amino acid sequence for the p15E-coding region of strain 129 G_{IX}^+ cDNA clone L510 are compared with the corresponding sequences of cDNA clones E2 and S13 and of infectious viruses MCF 247, Mo-MCF, and AKV as described in the legend to Fig. 2. The first nucleotide is one nucleotide 3′ to the last nucleotide in Fig. 4. The wavy line indicates a highly conserved, potentially transmembrane domain. The arrow indicates the predicted cleavage site which generates the R-peptide.

and BALB/c) and were derived from distinct ecotropic parental viruses (AKV and Mo-MuLV).

Sequence homology between endogenous retroviral *env* regions and recombinant MCF sequences. cDNA clones E1, L511, and L510 shared regions of sequence overlap, with 100% nucleotide identity over the sequenced regions. Because of this extensive homology, it is reasonable that the sequence presented reflects the sequence of a single, continuous RNA transcript. Two other independent clones from a portion of this region (E119 and L62) also shared 100% identity with the sequence presented over the 600 and 200 nucleotides, respectively, for which their sequences were determined (data not shown). This sequence, (Fig. 3 and 4) contained a 1,932-nucleotide open triplet reading frame which predicted a potential protein sequence of 644 amino acids representing the gp70-p15E polyprotein precursor.

gp70-coding region. The sequence of the strain 129 endogenous gp70-coding region more closely resembled the gp70-coding regions of recombinant viruses than those of ecotropic or xenotropic viruses. Two recombinant MCF viral

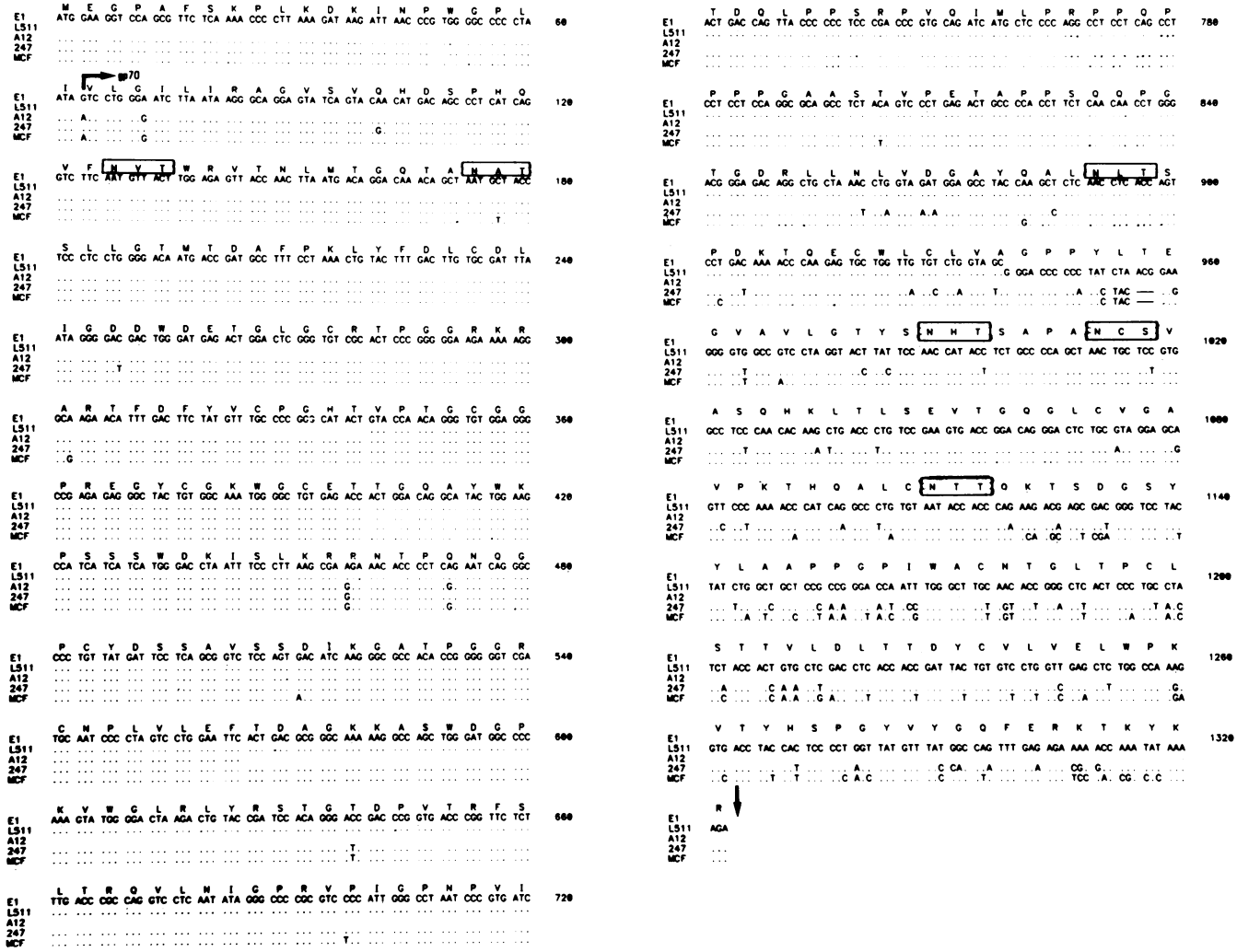

FIG. 4. Sequence comparison of the gp70-coding region. The DNA sequence and predicted amino acid sequence for the gp70-coding region of strain 129 G_{IX}^{+} endogenous transcripts were derived from two overlapping cDNA clones, E1 and L511. These sequences are compared with those of AKR endogenous proviral clone A12, MCF 247 virus, and Mo-MCF virus as described in the legend to Fig. 2. This sequence begins with nucleotide 303 (see Fig. 2) due to the overlapping, out-of-frame coding regions of *pol* and *env*. Mature gp70 protein is predicted to begin with the valine residue at positions 97 to 99 after cleavage from the 31-amino acid potential signal peptide. The boxes indicate potential glycosylation sites, and the arrow at the end of the sequence indicates the predicted site of cleavage from p15E.

sequences which showed considerable similarity to this RNA sequence are shown in Fig. 4. A similar high degree of nucleotide homology was found for two MCF viruses induced from C3H mouse cell lines (28), as well as for the sequences of two recombinant spleen focus-forming viruses, Friend spleen focus-forming virus (1) and Rauscher spleen focus-forming virus (4) (data not shown). However, to align the predicted endogenous gp70 sequence with the sequence of ecotropic Mo-MuLV (42), four insertions and five deletions would be required to juxtapose similar regions. Likewise, alignment with the available nucleotide sequences of NFS and NZB xenotropic gp70-coding regions (33, 37) would require insertion of four codons and deletion of one codon in the amino portion and deletion of one codon in the carboxy region. Although functional data are lacking for endogenously expressed *env* proteins, since they are not associated with viral particles, these structural comparisons classify them with the dualtropic or MCF family rather than ecotropic or xenotropic virus.

A sequence comparison of 1,323 nucleotides of the endogenous strain 129 gp70-coding region with the homologous regions of recombinant virus MCF 247 and Mo-MCF virus revealed that although there were 74 and 78 nucleotide differences, respectively, these differences were not randomly distributed but clustered toward the 3′ end of the sequence. Thus, the sequence of this transcribed *env* region is nearly identical to the corresponding 5′ two-thirds of the gp70-coding region of MCF viruses that is postulated to result from recombination of the ecotropic parent virus with endogenous genomic DNA. Within this conserved region only 4 of 859 nucleotides are different, resulting in three amino acid changes compared with MCF 247 virus. Similarly, when this strain 129 sequence was compared with the same region of the Mo-MCF genome, nine nucleotide differences were observed, seven of which led to changes in amino acids. The endogenous strain 129 gp70 sequence was also very similar to that of the endogenous proviral sequence, A12, which was cloned from the murine AKR

TABLE 1. Summary of homologous relationships between the strain 129 endogenous contiguous sequence and viral sequences[a]

Region	Nucleotide position	Sequence	No. of nucleotide differences within synonymous codons[b]	No. of nucleotide differences within different codons	Nucleotide difference (%)	Predicted amino acid difference (%)	No. of aligned nucleotides	No. of gaps introduced to create alignments
pol	1-360	E2	(15)[c]		4.8		313	1
		S13	(3)[c]		3.2		94	0
		A12	2	1	0.8	0.8	360	0
		MCF 247	2	1	0.8	0.8	360	0
		Mo-MCF	24	12	10.0	5.8	360	0
		AKV	33	20	15.2	10.3	348	1
gp70	1-801	A12	1	3	0.7	1.6	564	0
		MCF 247	1	3	0.5	1.1	801	0
		Mo-MCF	2	7	1.1	2.6	801	0
	802-1323	MCF 247	44	26	13.5	8.7	519	1
		Mo-MCF	29	40	13.3	12.7	519	1
p15E	1-543	E2	(3)[c]		5.1		59	0
		S13	(3)[c]		8.3		36	0
		MCF 247	75	33	20.0	10.0	540	1
		Mo-MCF	94	30	23.0	10.0	540	1
		AKV	76	34	20.1	10.0	540	1
R-peptide	544-606	E2	(3)[c]		4.8		63	0
		S13	(1)[c]		1.6		63	0
		MCF 247	3	1	6.3	4.8	63	0
		Mo-MCF	7	12	33.3	31.6	57	1
		AKV	6	7	22.8	21.1	57	1

[a] A contiguous sequence for the strain 129 endogenous transcript was assumed by using the longest sequences shown in Fig. 2 through 4.
[b] Number of nucleotide differences compared with the strain 129 sequence.
[c] Deleted endogenous sequences E2 and S13 do not contain open reading frames; therefore, the total numbers of nucleotide differences are reported.

genome (17) but is not known to be transcriptionally active. Of the nucleotide sequence available for this proviral element, only 4 differences among 564 nucleotides were observed. Interestingly, of these differences between the A12 and strain 129 sequences only one nucleotide difference at position 457 (Fig. 4) was shared with the MCF 247 virus sequence, indicating that this region of the MCF 247 genome, which was derived from the endogenous sequence, is more similar to the transcribed strain 129 sequence than to the AKR endogenous sequence. In comparison, alignment of the same region of the gp70-coding sequence of NZB xenotropic virus (33) demonstrated 8.4% nucleotide dissimilarity after allowing for a 12-nucleotide insertion and two 3-nucleotide deletions (data not shown).

The differences that do distinguish the nucleotide sequences of these endogenous transcripts from those of MCF viruses (Fig. 5) accumulate abruptly 3′ to the probable recombination point of MCF 247 at position 856 (Fig. 4) (14). These differences between the strain 129 sequence and the sequence of the ecotropic gp70-encoding region of MCF 247 account for 70 of the 74 overall nucleotide differences and result in 15 amino acid differences. Furthermore, in this portion of the *env* gene the introduction of a one-codon gap is required in the MCF virus sequence for alignment, resulting in the addition of a strain 129 transcript-specific threonine residue at positions 952 to 954. Similarly, 69 of the 78 nucleotide differences between the gp70-coding regions of the strain 129 transcripts and the Mo-MCF genome cluster at the 3′ end of the sequence, past the presumed site of recombination and within the ecotropic-derived region of Mo-MCF gp70, and also require the same threonine codon insertion for sequence alignment.

The comparisons described above (Table 1) demonstrate that there is 99% or greater sequence homology between the regions substituted in the generation of the recombinant viruses and the strain 129 transcribed sequence presented here. Therefore, the comparisons help define the region of ecotropic sequence that is not substituted in the gp70-coding region of MCF viruses and thus distinguished by 13.5% dissimilarity compared with the corresponding strain 129 sequence. The clustering of these differences is relevant to the mapping of potential sites of the recombinational events that presumably gave rise to Mo-MCF virus. Bosselman et al. (6) indicated that this recombination took place at position 1073 (Fig. 4) on the basis of similarity with Mo-MuLV. However, our data imply that recombination more likely took place further 5′, closer to the region implicated for MCF 247 recombination (position 856). Therefore, the region 3′ to this potential recombination point may actually be ecotropic in origin, the differences between Mo-MCF and Mo-MuLV in this region having been introduced by mutation rather than by recombination. Interestingly, of the differences between Mo-MCF and the strain 129 sequence, 31 are silent while 48 lead to codon changes; in MCF 247, 46 changes are silent while 28 lead to amino acid differences. This preponderance of coding changes may be indicative of a selective pressure favoring an altered protein sequence for Mo-MCF.

p15E-coding region. As shown in Fig. 3, the beginning of p15E is defined by analogy with Rauscher MuLV, with which it shares the amino-terminal protein sequence. The sequence shown is from clone L510, which contains the entire 606-nucleotide p15E open reading frame (202 amino acids, predicting a protein having a molecular weight of

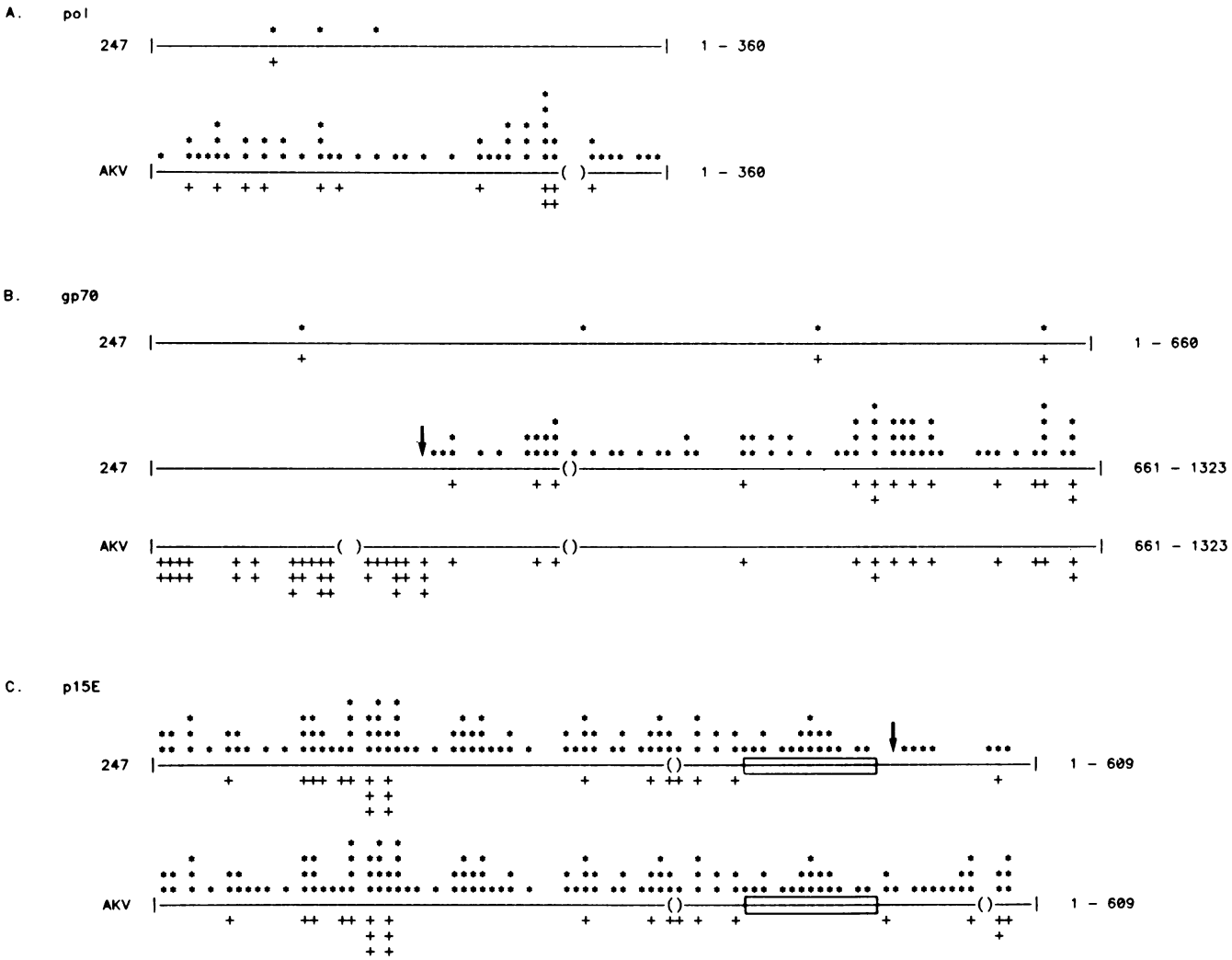

FIG. 5. Histogram display of sequence similarities. The histograms represent nucleotide and amino acid differences between the sequence indicated and the composite sequence derived for the strain 129 G_{IX}^+ endogenous transcript (see text). The lines represent identity with the strain 129 endogenous sequence, while asterisks above the line and plus signs below the line indicate the approximate positions of nucleotide and amino acid differences, respectively, in the sequence indicated at the left. Parentheses indicate gaps introduced in the sequence to increase alignment with the strain 129 endogenous sequence. (A) *pol* region from Fig. 2. (B) gp70 open reading frame from Fig. 4. (C) p15e open reading frame from Fig. 3. Since the first half of AKV gp70 is extremely dissimilar, only the second half of the gp70 region from AKV is compared in panel B. The arrows in panels B and C indicate two predicted recombination points for the generation of MCF 247 virus (see text). The boxes in panel C indicate the predicted transmembrane domain.

22,000). This clone overlaps L511, and, for the approximately 400 nucleotides determined, their sequences are identical. Furthermore, clone E119 also has the same sequence as L510 for approximately 350 nucleotides of the sequence determined (data not shown). However, the overlapping portion of clone S13 contains 4 differences out of 122 nucleotides, and that of clone E2 shows 6 differences out of 99 nucleotides 2 of which are shared with S13. Further sequence heterogeneity was detected in a p15E-coding region clone from thymus, which showed five nucleotide differences in the 3′ one-half of p15E, leading to two amino acid changes (Gln to Leu at positions 535 to 537 and Lys to Glu at positions 588 to 600).

The strain 129 endogenous transcript sequences encode a protein that is one amino acid longer than p15E from MCF 247 virus due to a codon insertion at positions 364 to 366, resulting in the addition of a serine residue. This endogenous transcript-specific codon was also present in clone E119 and in the thymus clone sequences. The endogenous transcript is three codons longer than the p15E region of Mo-MCF due to a second insertion (two codons at positions 580 to 585 [Asp-Pro]). This insertion is also present in MCF 247 virus, as well as in feline leukemia virus (9). Overall, MCF 247 virus has 112 nucleotide differences compared with the L510 sequence shown, 34 of them leading to changes in 19 amino acid residues. Similarly, Mo-MCF has 142 differences, with 38 leading to changes in 24 codons. These data are summarized in Table 1 for comparison of the relative differences in nucleotide and amino acid sequences. AKV is almost identical with MCF 247 virus over this region until the 3′ terminus, where MCF 247 virus is more similar to the strain 129 sequence, and was probably derived from a nonecotropic parent due to another recombinational event at about position 500 (16) (Fig. 1).

DISCUSSION

The sequences of actively transcribed, endogenous retrovirus-related RNA transcripts described here contain characteristic features demonstrating their close homology to infectious viruses, particularly recombinant MCF leukemogenic virus. In particular, the *env* region contains an open reading frame which potentially encodes a protein product closely resembling typical viral gp70 and p15E proteins, predicting that this RNA species is responsible for the serologically detected viral antigens of strain 129 mice (31, 45). This protein, derived from an open reading frame partially overlapping that of the *pol* gene, would be derived from a precursor containing a typical N-terminal signal peptide (5), allowing placement at the cell surface, and cleavage from p15E by a trypsin-like activity after the lysine-arginine doublet at positions 1318 to 1323 (Fig. 4). The mature predicted protein is 441 amino acids long and contains six attachment sites for N-linked glycosylation (29), all of which are conserved in MCF 247 and Mo-MCF gp70 proteins.

This endogenous *env* region transcript allows precise definition of the substituted sequences acquired by pathogenic MCF viruses. From these data, together with the partial nucleotide sequence of a genomic copy of the endogenous provirus-like sequence obtained from the AKR strain (17), it is clear that MCF 247 virus acquired an endogenous sequence similar to that constitutively expressed in strain 129 G_{IX}^+ mice for the 3′ portion of *pol*, the 5′ two-thirds of gp70, and the extreme 3′-terminal region of p15E while retaining ecotropic sequences for the remaining regions between these substitutions. Sequence comparisons in the U3 region indicate the presence of yet another substitution of nonecotropic information in the creation of MCF 247 virus (18, 35; D. E. Levy, R. McKinnon, J. W. Gautsch, and M. C. Wilson, manuscript in preparation).

The portion of the *pol* sequence described here is derived from the 3′ end of that gene, a region thought to encode a function necessary for the establishment of a productive viral infection (41). The major difference between ecotropic virus and the endogenous transcript within this region is the absence from ecotropic *pol* of four codons at the extreme 3′ end, accompanied by a clustering of amino acid differences (Fig. 2). Recombinant leukemogenic viruses share extensive homology with the endogenous transcript within this region. Interestingly, similar differences were observed between Mo-MuLV (42), which also exhibits leukemogenic activity, and nonpathogenic AKV (12). This portion of *pol* may contribute some leukemogenic potential to recombinant viruses (13).

The structural data reported here suggest that the strain 129 cellular gp70 proteins display dualtropic determinants similar to those of MCF viruses. The acquisition by MCF viruses of a similar endogenous sequence encoding the amino-terminal portion of gp70 is presumably responsible for their expanded host ranges compared with ecotropic viruses and may influence their tissue tropism (8). The carboxy-terminal portion of gp70 and the bulk of p15E differ from the corresponding regions of MCF viruses, which derive these coding sequences from the ecotropic parent. The reason for this return to the ecotropic sequence in MCF viruses is not clear, although this feature has been found to be a major determinant of leukemogenesis in recombinant viruses (13, 23, 27).

The nucleotide sequence of the endogenous transcript differs uniformly from the sequences of infectious ecotropic virus and MCF virus throughout the 3′ portion of the *env* gene (approximately 13 and 20% nucleotide differences [Table 1] for carboxy-terminal gp70 and p15E, respectively). However, the predicted amino acid sequence displays regions which have been highly conserved, punctuated by segments of clustered amino acid differences (Fig. 5). For example, the probable membrane-spanning, hydrophobic domain of ecotropic p15E (10, 22) is completely conserved in the endogenous protein at the amino acid level but exhibits 19 and 22% dissimilarity compared with MCF 247 virus and Mo-MCF virus at the nucleotide level (nucleotides 406 to 495) (Fig. 3). However, another hydrophobic domain at the amino terminus of AKV p15E, which may produce membrane-fusing activity (47), is disrupted by three charged residues (Asp, Lys, and Glu) in the endogenous protein, possibly resulting in altered activity. Regions in which the protein sequence has drifted compared with infectious virus may affect the interaction of endogenous gp70 and p15E and could result in the release of gp70 as a secreted form found in sera and epididymides of normal mice (11). Alternatively, the carboxy portion of gp70 could contribute to tropism determinants of potential viruses expressing the entire strain 129 endogenous gp70. Because of the availability of these cloned sequences we have begun to test the function of these regions directly by replacing the *env* components of gp70 and p15E, as well as the long terminal repeat of infectious virus, with the analogous regions of the endogenous strain 129 sequence.

The biological consequence of the normal expression of MCF-like envelope proteins in strain 129 mice may be the blockage of the specific MCF cell surface receptors (36), leading to resistance to MCF viral infection (2). Since the MCF viruses are strongly implicated in the incidence of leukemia, such resistance may impart a significant selective advantage to the animal (39). The close sequence similarity between these RNA species and MCF virus also suggests a relationship between active expression of the proviruses and the generation of leukemogenic viruses. It is interesting that of the many proviral sequences of the AKR genome, the one found to be most similar to the substituted region of MCF 247 virus (17) is also highly homologous to the expressed copy in strain 129. Furthermore, the large *env* deletions detected in clones S13 and E2 are strikingly similar to the substituted region of MCF virus. The 5′ breakpoint of S13 corresponds exactly to the 5′ recombination point identified for MCF virus CI-3 (28), while the 3′ breakpoint of E2 matches the recombination point in the p15E region of MCF 247 virus (16).

ACKNOWLEDGMENTS

We thank R. Ogata and E. Rothenberg for helpful discussions; J. Elder, J. G. Sutcliffe, and P. Policastro for comments on the manuscript; and P. Graber and A. McDonald for expert help in manuscript preparation.

This work was supported in part by Public Health Service Predoctoral Training Grant GM-07616 from the National Institutes of Health to D.E.L. and by Public Health Service grants CA-27489 and CA-33730 from the National Institutes of Health to R.A.L. and M.C.W., respectively.

LITERATURE CITED

1. **Amanuma, H., A. Katori, M. Obata, N. Sagata, and Y. Ikawa.** 1983. Complete nucleotide sequence of the gene for the specific glycoprotein (gp55) of Friend spleen focus-forming virus. Proc. Natl. Acad. Sci. USA **80**:3913–3917.
2. **Bassin, R. H., S. Ruscetti, I. Ali, D. Haapala, and A. Rein.** 1982.

Normal DBA/2 mouse cells synthesize a glycoprotein which interferes with MCF virus infection. Virology **123**:139–151.

3. **Berns, A. J., M. H. T. Lai, R. A. Bosselman, M. A. McKennett, L. T. Bacheler, H. Fan, E. C. Robanus Maandag, H. van der Putten, and I. M. Verma.** 1980. Molecular cloning of unintegrated and a portion of integrated Moloney murine leukemia viral DNA in bacteriophage lambda. J. Virol. **36**:254–263.
4. **Bestwick, R. K., B. A. Boswell, and D. Kabat.** 1984. Molecular cloning of biologically active Rauscher spleen focus-forming virus and the sequences of its *env* gene and long terminal repeat. J. Virol. **51**:695–705.
5. **Blobel, G., and B. Dobberstein.** 1975. Transfer of proteins across membranes. J. Cell Biol. **67**:835–851.
6. **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 proviral DNA. J. Virol. **44**:19–31.
7. **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.
8. **Devare, S. G., U. R. Rapp, G. J. Todaro, and J. R. Stephenson.** 1978. Acquisition of oncogenicity by endogenous mouse type C viruses: effects of variations in *env* and *gag* genes. J. Virol. **28**:457–465.
9. **Elder, J. H., and J. I. Mullins.** 1983. 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. J. Virol. **46**:871–880.
10. **Green, N., T. M. 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. USA **78**:6023–6027.
11. **Hara, I., I. Shozo, and F. J. Dixon.** 1982. Murine serum glycoprotein gp70 behaves as an acute phase reactant. J. Exp. Med. **155**:345–357.
12. **Herr, W.** 1984. Nucleotide sequence of AKV murine leukemia virus. J. Virol. **49**:471–478.
13. **Holland, C. A., J. W. Hartley, W. P. Rowe, and N. Hopkins.** 1985. At least four viral genes contribute to the leukemogenicity of murine retrovirus MCF 247 in AKR mice. J. Virol. **53**:158–165.
14. **Holland, C. A., J. Wozney, and N. Hopkins.** 1983. Nucleotide sequence of the gp70 gene of murine retrovirus MCF 247. J. Virol. **47**:413–420.
15. **Horowitz, J., and R. Risser.** 1982. A locus that enhances the induction of endogenous ecotropic murine leukemia viruses is distinct from genome-length ecotropic proviruses. J. Virol. **44**:950–957.
16. **Kelley, 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.
17. **Khan, A. S.** 1984. Nucleotide sequence analysis establishes the role of endogenous murine leukemia virus DNA segments in formation of recombinant mink cell focus-forming murine leukemia viruses. J. Virol. **50**:864–871.
18. **Kahn, A. S., and M. A. Martin.** 1983. Endogenous murine leukemia proviral long terminal repeats contain a unique 190-base-pair insert. Proc. Natl. Acad. Sci. USA **80**:2699–2703.
19. **Khan, A. S., W. P. Rowe, and M. A. Martin.** 1982. Cloning of endogenous murine leukemia virus-related sequences from chromosomal DNA of BALB/c and AKR/J mice: identification of an *env* progenitor of AKR-247 mink cell focus-forming proviral DNA. J. Virol. **44**:625–636.
20. **Kozak, C. A., and W. P. Rowe.** 1980. Genetic mapping of xenotropic murine leukemia virus-inducing loci in five mouse strains. J. Exp. Med. **152**:1419–1423.
21. **Kozak, C. A., and W. P. Rowe.** 1982. Genetic mapping of ecotropic murine leukemia virus-inducing loci in six inbred strains. J. Exp. Med. **155**:524–534.
22. **Lenz, J., R. Crowther, A. Straceski, and W. Haseltine.** 1982. Nucleotide sequence of the Akv *env* gene. J. Virol. **42**:519–529.
23. **Lenz, J., and W. A. Haseltine.** 1983. Localization of the leukemogenic determinants of SL3-3, an ecotropic, XC-positive murine leukemia virus of AKR mouse origin. J. Virol. **47**:317–328.
24. **Levy, D. E., R. A. Lerner, and M. C. Wilson.** 1982. A genetic locus regulates the expression of tissue specific mRNAs from multiple transcription units. Proc. Natl. Acad. Sci. USA **79**:5823–5827.
25. **Levy, D. E., R. A. Lerner, and M. C. Wilson.** 1985. The Gv-1 locus coordinately regulates the expression of multiple endogenous murine retroviruses. Cell **41**:289–299.
26. **Levy, J. A.** 1978. Xenotropic type C viruses. Curr. Top. Microbiol. Immunol. **79**:109–213.
27. **Lung, M. L., J. W. Hartley, W. P. Rowe, and N. H. Hopkins.** 1983. Large RNase T1-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.
28. **Mark, G. E., and U. R. Rapp.** 1984. Envelope gene sequence of two in vitro-generated mink cell focus-forming murine leukemia viruses which contain the entire gp70 sequence of the endogenous nonecotropic parent. J. Virol. **49**:530–539.
29. **Marshall, R. D.** 1974. The nature and metabolism of the carbohydrate-protein linkages of glycoproteins. Biochem. Soc. Symp. **40**:17–26.
30. **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.
31. **Obata, Y., H. Ikeda, E. Stockert, and E. A. Boyse.** 1975. Relation of G_{IX} antigen of thymocytes to envelop glycoprotein of murine leukemia virus. J. Exp. Med. **141**:188–197.
32. **Old, L. J., and E. Stockert.** 1977. Immunogenetics of cell surface antigens of mouse leukemia. Annu. Rev. Genet. **11**:127–160.
33. **O'Neill, R. R., C. E. Buckler, T. S. Theodore, M. A. Martin, and R. Repaske.** 1985. Envelope and long terminal repeat sequences of a cloned infectious NZB xenotropic murine leukemia virus. J. Virol. **53**:100–106.
34. **Orcutt, B. C., M. O. Dayhoff, and W. C. Barker.** 1982. User's guide for the alignment score program. Report 820501-08710. National Biomedical Research Foundation, Washington, D.C.
35. **Quint, W., W. Boelens, P. van Wezenbeek, T. Cuypers, E. R. Maandag, G. Selten, and A. Berns.** 1984. Generation of AKR mink cell focus-forming viruses: a conserved single-copy xenotropic-like provirus provides recombinant long terminal repeat sequences. J. Virol. **50**:432–438.
36. **Rein, A., and A. Schultz.** 1984. Different recombinant murine leukemia viruses use different cell surface receptors. Virology **136**:144–152.
37. **Repaske, R., R. R. O'Neill, A. S. Khan, and M. A. Martin.** 1983. Nucleotide sequence of the *env*-specific segment of NFS-TH-1 xenotropic murine leukemia virus. J. Virol. **46**:204–211.
38. **Risser, R., J. M. Horowitz, and J. McCubrey.** 1983. Endogenous mouse leukemia viruses. Annu. Rev. Genet. **17**:85–121.
39. **Ruscetti, S., L. Davis, J. Feild, and A. Oliff.** 1981. Friend murine leukemia virus-induced leukemia is associated with the formation of mink cell focus-inducing viruses and is blocked in mice expressing endogenous mink cell focus-inducing xenotropic viral envelope genes. J. Exp. Med. **154**:907–920.
40. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5476.
41. **Schwartzberg, P., J. Colicelli, and S. P. Goff.** 1984. Construction and analysis of deletion mutations in the *pol* gene of Moloney murine leukemia virus: a new viral function required for productive infection. Cell **37**:1043–1052.
42. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukaemia virus. Nature (London) **293**:543–548.
43. **Stockert, E., E. A. Boyse, Y. Obata, H. Ikeda, N. H. Sarker, and**

H. A. Hoffman. 1975. New mutant and congenic mouse stocks expressing the murine leukemia virus-associated thymocyte surface antigen Gix. J. Exp. Med. **142:**512–517.

44. **Stockert, E., L. J. Old, and E. A. Boyse.** 1971. The G_{IX} system. A cell surface allo-antigen associated with murine leukemia virus; implications regarding chromosomal integration of the viral genome. J. Exp. Med. **149:**200–215.
45. **Strand, M., F. Lilly, and J. T. August.** 1974. Host control of endogenous murine leukemia virus gene expression: concentrations of viral proteins in high and low leukemia mouse strains. Proc. Natl. Acad. Sci. USA **71:**3682–3686.
46. **Varmus, H. E.** 1984. The molecular genetics of cellular oncogenes. Annu. Rev. Genet. **18:**553–612.
47. **White, J., M. Kielian, and A. Helenius.** 1983. Membrane fusion proteins of enveloped animal viruses. Q. Rev. Biophys. **16:**151–195.