JOURNAL OF VIROLOGY, Feb. 1984, p. 530–539
0022-538X/84/020530-10$02.00/0

Vol. 49, No. 2

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

GEORGE E. MARK†* AND ULF R. RAPP

Laboratory of Viral Carcinogenesis, National Cancer Institute, Frederick, Maryland 21701

Received 13 June 1983/Accepted 4 October 1983

The mink cell focus-forming (MCF) class of recombinant murine leukemia viruses (CI-1 to 4) were isolated from iododeoxyuridine-induced C3H/MCA 5 cells in culture and molecularly cloned. These genomes included infectious (CI-3) and defective (CI-4) recombinants. A total of 2,408 nucleotides of CI-3 virus DNA, including the MCF envelope gene, were sequenced and compared with ecotropic, dual-tropic, and xenotropic sequences. The extent of recombinational exchange in CI-3 was from 145 nucleotides 3′ of the splice acceptor site for the envelope mRNA to nucleotide 1,722, between the end of gp70 and the beginning of Prp15E. Thus, the entire gp70 sequence of the endogenous nonecotropic parent was present in this recombinant. The nature and location of the recombinant junctions were consistent with a mechanism involving DNA exchange during reverse transcription. Comparison of the substituted sequence in CI-3 with that of Moloney MCF virus suggests a very close relationship, if not identity, between the endogenous dual-tropic proviruses from which they were derived. A nonidentity of xenotropic and MCF gp70s was observed, suggesting that xenotropic murine leukemia viruses are not the nonecotropic parent of the *env* gene of MCF murine leukemia viruses. The replication-defective virus CI-4 had a 684-nucleotide deletion present in the *env* gene, eliminating the hydrophobic regions within the gp70 carboxy end and the p15E amino end. This sequence was bordered by an 11-nucleotide direct repeat in CI-3 viral DNA.

Mink cell focus-forming murine leukemia viruses (MCF-MuLVs) are recombinants between ecotropic and endogenous xenotropic related MuLV (2, 4, 8, 31) sequences. The recombination occurs in vivo during the preleukemic period in high-leukemic strains of mice (31) and affects the envelope gene (2, 4, 8, 19) as well as the long terminal repeat (LTR) of the ecotropic parent virus (13, 38). The envelope gene substitution introduces a gp70 with a new receptor binding site (28) which, perhaps together with new LTR sequences (13, 38), alters the viral host range in vitro from eco- to amphotropic and might also affect its tissue tropism in vivo (7). Some MCF viruses accelerate leukemia development upon intrathymic inoculation into young mice of their strain of origin. The resulting tumors are monoclonal (11) and contain MCF provirus(es) integrated in specific chromosomal regions (40). It therefore appears likely that the final step in transformation by MCF-MuLV involves insertional mutagenesis in specific target cells, such that a resident oncogene is activated.

We have reported previously the isolation and molecular cloning (27) of MCF-MuLV from cultures of iododeoxyuridine-induced C3H/10T1/2 cells. One of the molecular clones, CI-3, induced anchorage-independent growth of epithelial mink lung cells in vitro (27). This latter activity required the presence of nontransforming helper MuLV and epidermal growth factor (27). Unintegrated DNA intermediates present in cultures of in vitro-transformed epithelial mink lung cells were also the source of the molecular clone CI-4, a variant form of CI-3 virus, which contained a deletion in the p15E region of the envelope gene as does spleen focus-forming virus (SFFV) (17).

* Corresponding author.
† Present address: Laboratory of Human Carcinogenesis, National Cancer Institute, Bethesda, MD 20205.

The altered *env* gene products of MCF-MuLVs have been implicated as determinants of pathogenicity for both thymoma-inducing (9, 32) and erythroleukemia-inducing isolates (17). Proposed mechanisms include (i) a mitogenic activity of these envelope gene products mediated through their binding to specific immune receptors (22), or as a consequence of their cross-reactivity with normal tissue-specific growth factor receptors (26); (ii) lack of superinfection interference in sensitive target cells due to sluggish processing of the gp70 precursor (Prgp85), which ultimately leads to the accumulation of proviral DNA available for integration into sensitive sites (28, 37); or (iii) growth stimulation of infected immune cells by the host's immune response to viral envelope antigens present on their cell surface (15).

The viral envelope gene codes for two membrane proteins, gp70 and p15E, which are proteolytically cleaved from a glycosylated precursor (Prgp85) (10). Whereas both ecotropic MuLV envelope gene products have been obtained and partially sequenced (23), little information is available for MCF viral envelope proteins. To determine the exact nature of the in vitro-acquired *env* sequences in CI-3 and to more precisely define the extent of recombinational exchange in CI-3 as well as the deletion in CI-4, we have determined the envelope gene sequence from CI-3 and CI-4 and compared it with that of ecotropic AKR-MuLV (16). The MCF-specific substitution present in CI-3 was also compared with that of Moloney-MCF (Mo-MCF) (1), a pathogenic recombinant MuLV derived from a BALB/Mo mouse.

MATERIALS AND METHODS

Plasmid DNA preparation. The isolation, molecular cloning, and subcloning in pBR322 of the MCF-MuLV genomes pCI-2, pCI-3, and pCI-4 from in vitro-transformed epithelial mink lung cells (Mv-1Lu) have been previously described (27). Plasmid DNAs, prepared by the procedure of Clewell

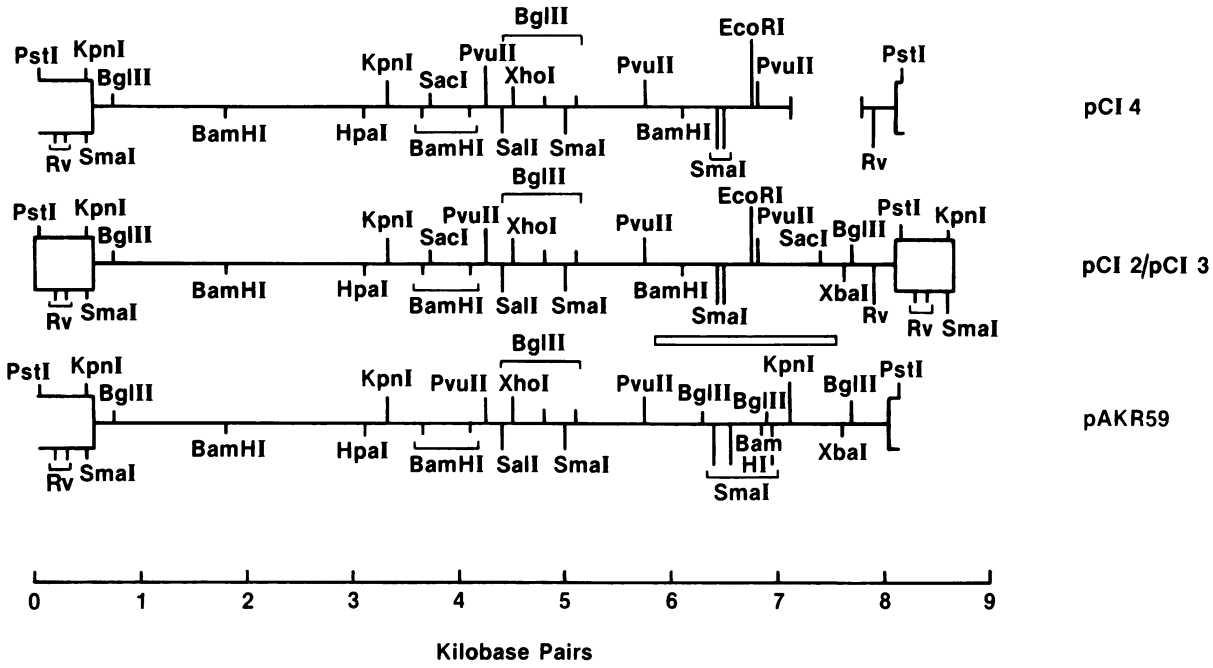


FIG. 1. Partial restriction endonuclease maps of the pCI and AKV DNA clones. The proviruses are oriented such that the 5′ ends are at the left and the 3′ ends are at the right. Large boxes or partial open boxes represent LTR sequences. Clone pCI-4 contains a deletion in its *env* gene shown by the gap in the horizontal line. The MCF-MuLV-cloned genomes pCI-2 and pCI-3 differ in restriction endonuclease sites from those of an ecotropic AKV (pAKR59) genome as indicated by the boxed area between the respective line drawings. The restriction enzymes shown are written in their entirety except *Eco*RV is abbreviated as Rv.

and Helinski (5), were purified by two successive cesium chloride isopycnic centrifugations. After treatment with proteinase K (Beckman) at 100 μg/ml for 60 min at 37°C, the DNAs were phenol extracted and concentrated by ethanol precipitation.

DNA sequence determination and analysis. DNA (~10 μg) to be sequenced was digested with the appropriate restriction endonuclease (New England Biolabs or Bethesda Research Laboratories) under conditions recommended by the manufacturer. To expose the 5′-phosphates of the *Stu*I- and *Sma*I-cleaved DNAs, their 3′ ends were resected by incubation at 37°C for 30 min with 0.15 U of *Exo*III nuclease (Miles Laboratories) in 0.05 M Tris-hydrochloride (pH 8.1)–0.005 M $MgCl_2$–0.01 M 2-mercaptoethanol. Bacteriophage T4 polynucleotide kinase (Bethesda Research Laboratories) and [γ-^{32}P]ATP (>5,000 Ci/mmol; Amersham Corp.) were used to label fragments with extended 5′ ends, subsequent to their dephosphorylation with bacterial alkaline phosphatase (Bethesda Research Laboratories), as described by Maxam and Gilbert (21). Terminal deoxynucleotide transferase (Boehringer Mannheim) and [α-^{32}P]ATP (3,000 Ci/mmol; Amersham Corp.) were used to label fragments with extended 3′ ends (21). DNA fragments containing uniquely labeled ends were isolated from Seaplaque (Seakem) agarose gels (6). The chemical modification method of DNA sequencing described by Maxam and Gilbert (21) was used except the guanine-plus-adenine reaction was carried out by incubation at 20°C for 8 min in 62.5% formic acid and then stopped with HZ stop buffer (21). Products were separated in 8 or 20% polyacrylamide–8 M urea sequencing gels, 0.4 mm thick, and visualized by autoradiography, using Kodak XAR-5 film at −70°C. DNA sequences were compiled and analyzed for protein sequence by the program of Queen and Korn (25).

RESULTS

Localization of the recombinant *env* gene sequences. Restriction maps of the viral genomes used for DNA sequencing are shown in Fig. 1 compared with that of AKV clone pAKR59, a subclone of AKR-623 (18), whose ecotropic *env* sequence has been determined by Lenz et al. (16). This comparison, as well as previous heteroduplex analyses (27), suggests that the recombinational borders were in the vicinity of the *Bam*HI site at 6.1 kilobases and the *Sac*I site at 7.4 kilobases. Although apparently complete viral genomes had been cloned (pCI-2 and pCI-3), only pCI-3 DNA produced infectious virus upon transfection of NIH 3T3 cells (27). Thus, this DNA was chosen as the subject of nucleotide sequence analysis.

DNA sequence analysis. All DNA sequences were determined by the method of Maxam and Gilbert (21). The strategy used to obtain the nucleotide sequence of the recombinant *env* gene of clone pCI-3 DNA is shown in Fig. 2. The sequence of 2,408 nucleotides, generated by this approach, contained 473 nucleotides of the 3′ end of the *pol* gene, an open reading frame of 1,917 nucleotides, representing the entire *env* gene, and sequences which stretch 39 bases into the U3 region of the LTR (Fig. 3). To determine the location of the recombinant junctions, this sequence was compared with corresponding sequences obtained by Lenz et al. (16) for AKR-MuLV. Although these authors do not present extensive sequence information 5′ of the ecotropic *env* gene, a comparison (Fig. 4A) suggests that the 5′ end of the recombinant *env* gene in pCI-3 DNA lies just upstream from the *Bam*HI site located at nucleotides 349 to 354. A recently obtained sequence of this region in MCF-247 (Nancy Hopkins, personal communication) compared with pCI-3

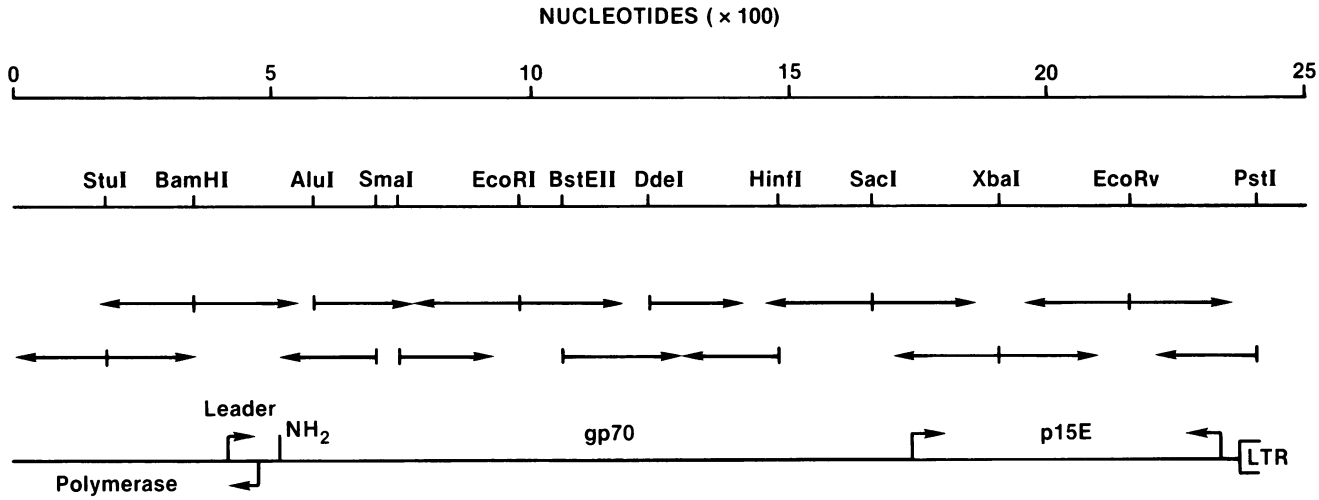


FIG. 2. Strategy for sequencing cloned pCI-3 DNA. Arrows show the direction of sequencing from a uniquely labeled end, indicated by a horizontal line. Boundaries of the polymerase and envelope genes and LTR sequences were determined from the sequence. Nucleotide numbering was derived from the sequence presented in Fig. 3.

reveals extensive base pair mismatch 5′ from this site, consistent with the interpretation that these sequences are endogenous MCF parental in MCF-247 and ecotropic parental in pCI-3. Thus, the substitution began in the carboxy terminus of the *pol* gene, approximately 130 nucleotides before the initiator AUG of the pCI-3 *env* gene and 140 to 150 nucleotides after the *env* splice acceptor sequence (nucleotides 126 to 135; Fig. 3). The same region was reported by Bosselman et al. (1) to be the beginning of a Mo-MCF recombinant *env* gene.

The 3′ limit of the pCI-3 *env* gene substitution appears to be 14 nucleotides before the amino terminus of p15E (Fig. 4B). Comparison of the ecotropic pAKR59 sequences with those of pCI-3 DNA show extensive mismatch until position 1,722 in CI-3, after which there are only six base changes in the remaining 687 nucleotides sequenced. In addition to base pair substitutions, pCI-3 DNA contains a three-nucleotide insertion at the 3′ junction between MCF and ecotropic sequences. The noninfectious pCI-2 DNA also contains an insertion at the same location; however, significantly more nucleotides are involved. A translation frameshift would occur as a consequence of the pCI-2 DNA sequence, resulting in the termination of polypeptide synthesis five codons into the p15E coding region. This would account for the noninfectious nature of the pCI-2 genome.

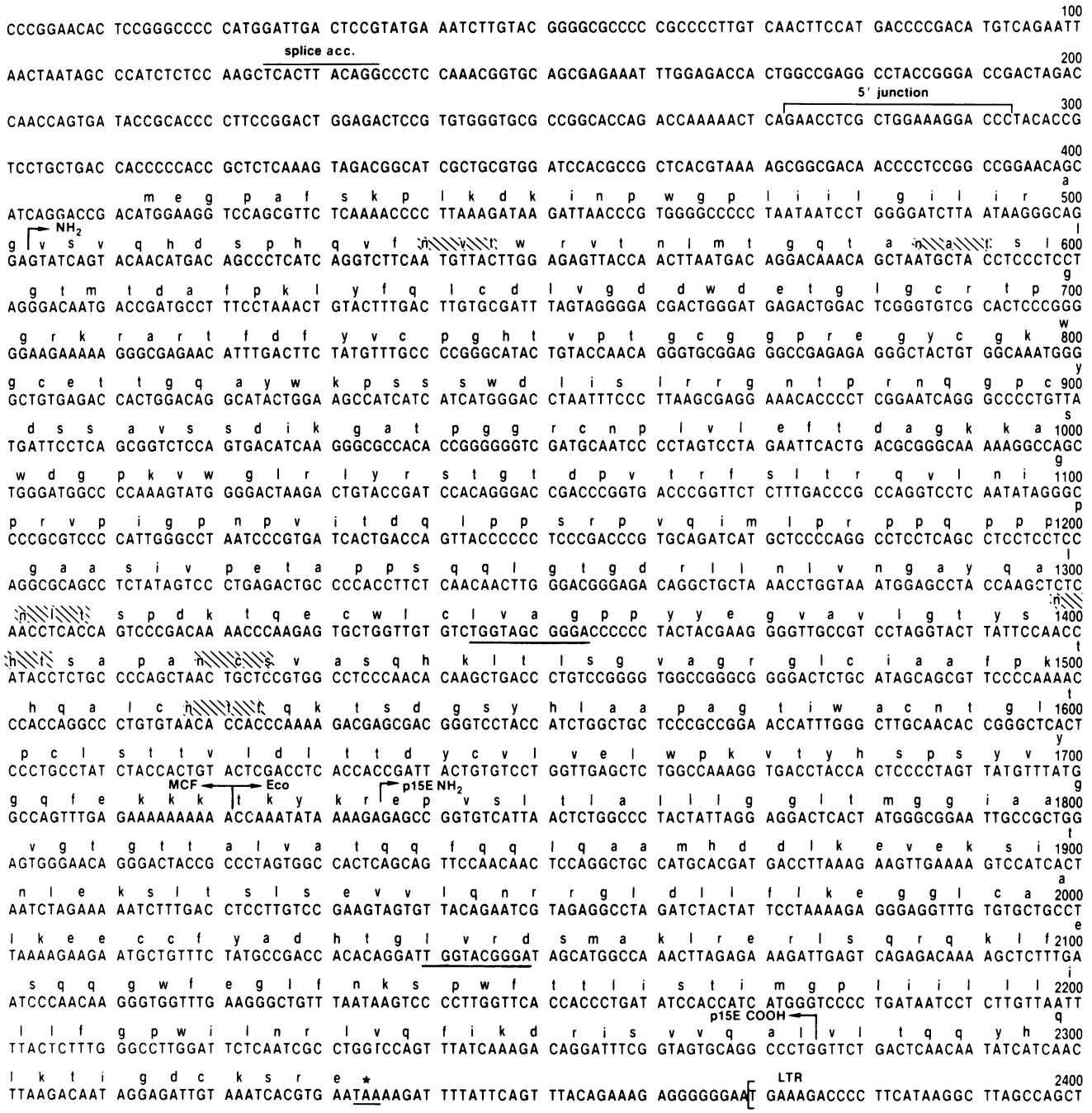
Thus, the extent of recombinant sequences in pCI-3 DNA spans 1,400+ nucleotides from a position near nucleotide 300 to position 1,722, encompassing the entire gp70 portion of the *env* gene. The discrepancy between this finding and heteroduplex observations which underestimated the size of the substitution is largely the result of the presence of a guanine plus cytosine-rich region followed by highly conserved ecotropically related sequences (84% nucleotide homology) from nucleotide 1,262 to the 3′ recombinant junction (those sequences defined as "Eco" in Fig. 5 and 7).

Prior electron microscopic analysis of pCI-3 DNA duplexed with AKV DNA had revealed two major regions of nonhomology separated by approximately 130 nucleotides of partial homology (27). These conserved sequences were found, as predicted, more than 1,600 nucleotides 5′ of the terminal *Pst*I site (Fig. 4C and 5). The homology of this region degenerates toward its central core. An identical region of partial homology is also found in the Moloney (35) and Friend (3, 14) ecotropic envelope genes, as well as in NFS-Th-1, a xenotropic virus (30).

The nucleotide sequence of Prp15E pCI-3 reveals only a 1% divergence from the AKR-MuLV ecotropic sequence (a difference of 6 of 600 nucleotides) and a change in two amino acids (Fig. 7). Missing in the pCI-3 DNA sequence are the six base insertions and five base substitutions described as unique to the leukemogenic MCF-247 Prp15E coding region, as well as the nucleotide changes responsible for the generation of the "MCF-specific" T1 oligonucleotide 101 (13). Of the 73 nucleotides sequenced beyond the Prp15E termination codon, including the plus-strand initiation region and the proximal U3 LTR region, the homology is 100 and 93% with respect to AKV and MCF-247, respectively. Thus, pCI-3 does not contain the additional sequences of MCF-247 obtained from some as yet to be defined nonecotropic parent.

Predicted amino acid sequence. The nucleotide sequence of pCI-3 predicts the *env* gene product to be 640 amino acids long (Fig. 7). This is 29 amino acids shorter than the sequence predicted for the ecotropic *env* gene of AKV (16) and 1 amino acid longer than that predicted for the Mo-MCF *env* gene (35). The first AUG codon (nucleotides 413 to 415; Fig. 3) is found 279 nucleotides downstream from the putative *env* gene splice acceptor sequence (nucleotides 126 to 135). It is this methionine which initiates the only long open reading frame dictated by the DNA sequence. The 30-amino acid signal peptide, cleaved from the N terminus of the primary gene product, is 1 amino acid shorter than the equivalent ecotropic peptide and exhibits only a 50% homology with the latter sequence. The location of this and other post-translational processing sites which define the amino and carboxy termini of the *env* gene proteins has been deduced by comparison with the known amino acid sequences of Mo-MuLV (35), Rauscher MuLV (23, 33), and Friend MuLV (14, 23).

The MCF-gp70 domain contains six canonical sequences, Asn-X-Thr or Asn-X-Ser (20), which can serve as sites of *N*-glycosylation (underlined in Fig. 7; CHO in Fig. 5). An



AACTGCAG

FIG. 3. Nucleotide sequence and deduced amino acid sequence determined for the pCI-3 *env* gene. Number at the right of each line refer to the number of nucleotides from the beginning of the presented sequence. Lowercase letters above the sequence show the predicted amino acid sequence. Also indicated are the location of: the *env* gene splice acceptor site (splice acc.); the presumed 5' recombinant junction (overlined); post-translational cleavage sites; potential glycosylation sites (shaded boxes); the presumed 3' recombinant junction (MCF <- -> Eco); and the 11-base pair direct repeats which define the deleted regions of pC-4 (underlined).

additional site of possible *N*-glycosylation is located in the p15E sequence (amino acids 574 to 576). This site does not seem to be glycosylated, since p15E cannot be labeled with radioactive glucosamine (24). Five of the six potential *N*-glycosylation sites are shared with the presumptive ecotropic AKV-MuLV (16) parent (Fig. 5) as well as Friend MuLV (14, 23) and ecotropic Mo-MuLV (35) *env* gene sequences. Conservation of additional ecotropic gp70 polypeptide features are evident from the deduced MCF gp70 protein sequence (Fig. 5) and will be referred to as "Eco"-like. The overall amino acid homology to AKR of this Eco-like domain is 87%. This is in contrast to the amino-terminal

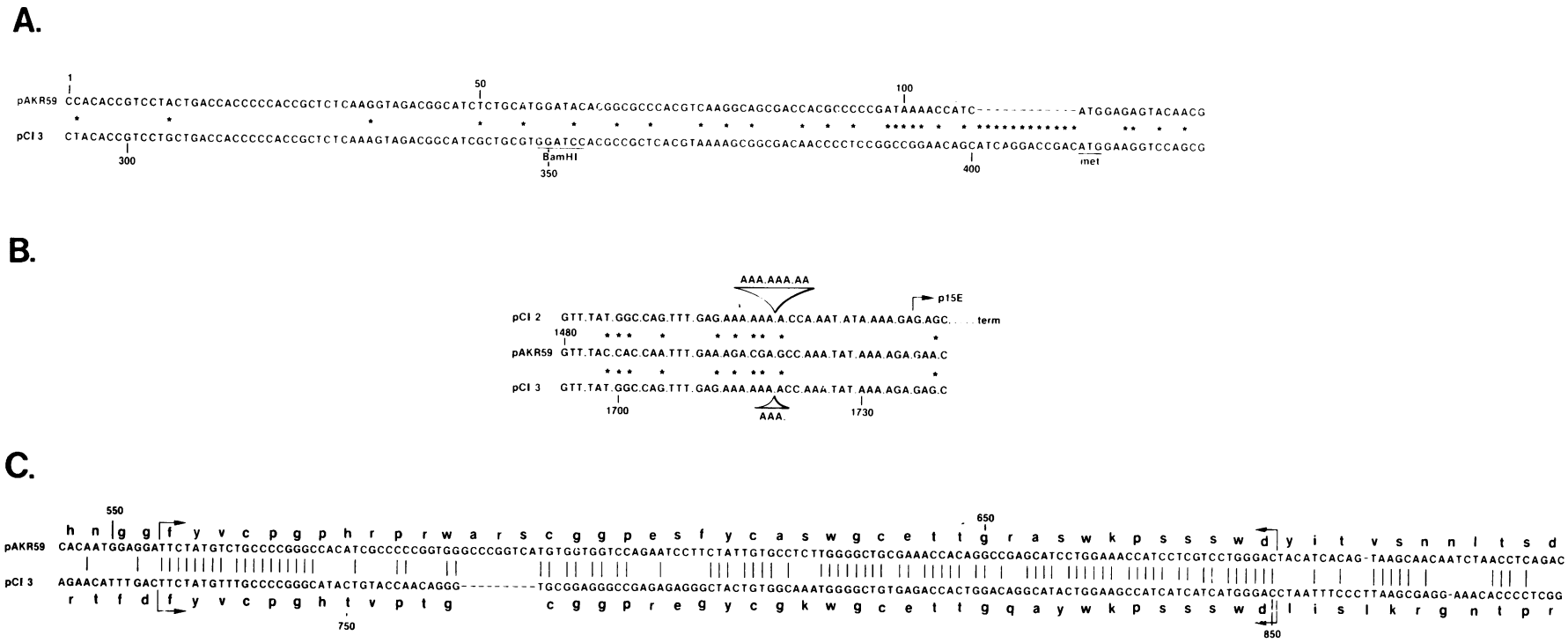


FIG. 4. Comparison of pCI sequences with ecotropic AKV (pAKR59). pCI nucleotide sequences are numbered as in Fig. 3, whereas pAKR59 sequences are numbered as described by Lenz et al. (16). Where present, astericks denote nonhomologous nucleotides and vertical lines denote homologous nucleotides; dashes represent deletions; and the standard one-letter amino acid code shows the predicted amino acid sequence. (A) Comparison of nucleotide sequences near the 5′ recombinant junction. Met defines the methionine initiator codon of the *env* gene. (B) Comparison of nucleotide sequences surrounding the 3′ recombinant junction. Triangles indicate the location of sequences inserted relative to the ecotropic sequence. The gp70/p15E cleavage site is indicated with an arrow. Codon reading frame is shown with periods. (C) Comparison of nucleotide and amino acid sequences surrounding the region of partial homology. The region of partial homology lies within the sequences bracketed by arrows.

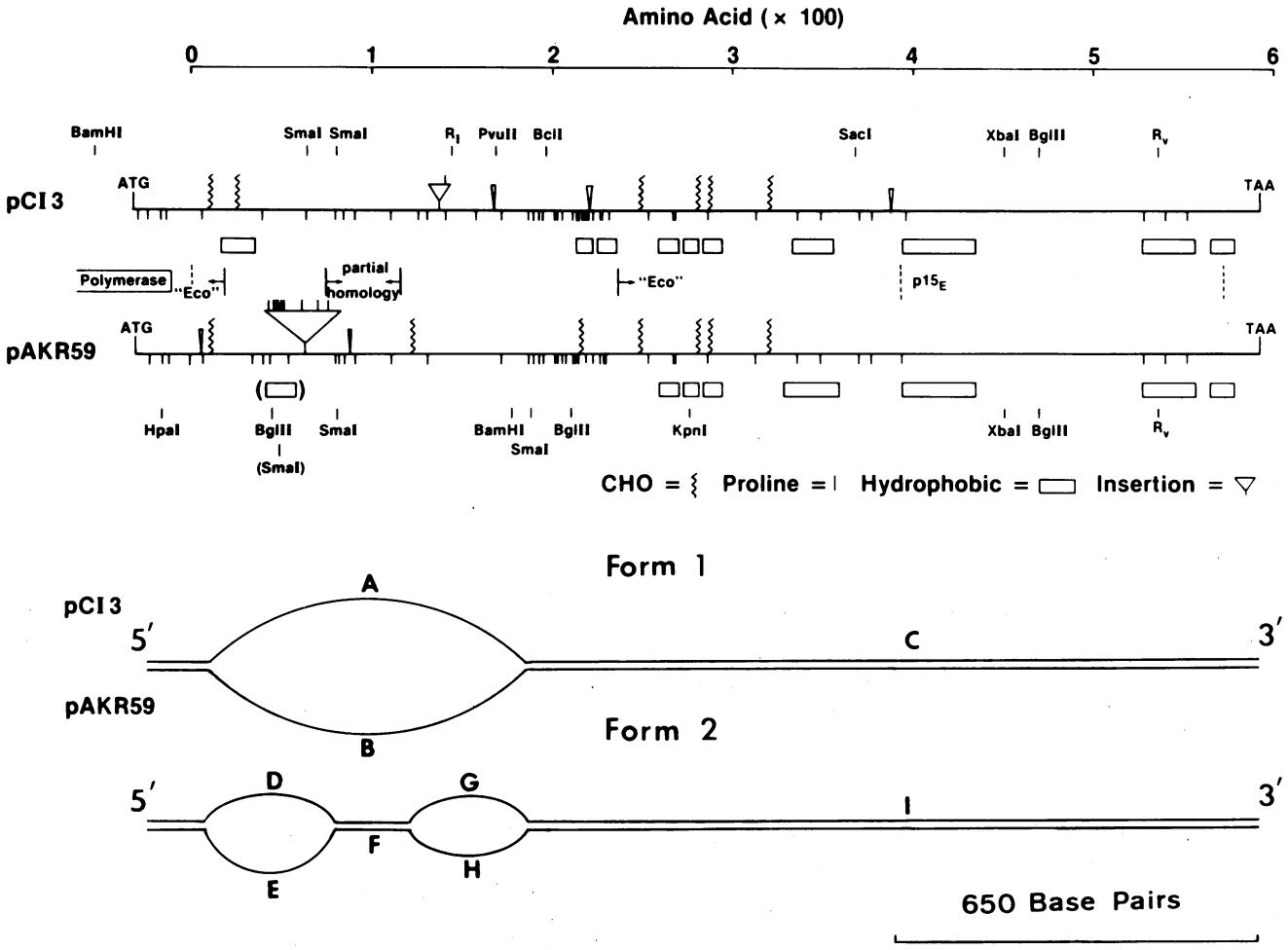


FIG. 5. Schematic comparisons of the MCF envelope sequence of pCI-3 and the ecotropic envelope sequence of AKV. Amino acid sequences were aligned so as to maximize structural similarities. Post-translational cleavage sites are shown as vertical dashed lines; characteristics within inserted regions are within parentheses; and ecotropic-like sequences are designated "Eco." Restriction endonuclease sites are shown (RI, *Eco*RI; RV, *Eco*RV), as are the location of glycosylation sites (CHO), proline residues, and extensive hydrophobic regions. Amino acid numbering begins at the mature NH_2 terminus of Prgp85 and is not influenced by the presence of inserted sequences. Below are interpretive drawings of the two forms of pCI-3/ARV heteroduplexed DNA seen previously (27). Contour lengths (in kilobases) were as follows: A, 0.69 ± 0.06; B, 0.83 ± 0.12; C, 1.32 ± 0.09; D, 0.32 ± 0.03; E, 0.49 ± 0.04; F, 0.13 ± 0.02, G, 0.25 ± 0.02; H, 0.26 ± 0.03; I, 1.28 ± 0.07.

portion, which exhibits only small sequence homologies with its ecotropic relatives. Adjacent to the Eco-like carboxy-terminal half of the MCF gp70, the position of apolar stretches and proline and cysteine (not shown) residues are conserved. These latter features may be dictated by structural requirements intrinsic to *env* glycoproteins (16). The Prp15E portion of the pCI-3 DNA was not substantially altered (197 identities of 199 amino acids) compared with the AKV-MuLV sequence (Fig. 7). Unlike MCF-247, where recombination altered the carboxy terminus to -Ser-Ile-Asp-Pro-Glu-Glu-Val-Glu-Ser-Arg-Glu-COOH, pCI-3 contained a two-amino acid-shortened sequence (-Thr-Ile-Gly-Asp-Cys-Lys-Ser-Arg-Glu-COOH) which differs from that of AKV (16) only by a Glu—>Gly alteration (Fig. 7).

Location of the SFFV-like deletion in the defective MCF virus genome represented by clone pCI-4. One of the MCF virus genomes (CI-4) cloned from the Hirt DNA of an in vitro-transformed mink epithelial lung cell culture was defective for replication (27). It was found to contain a deletion judged by DNA heteroduplexing to be coincident with the *env* gene deletion in Friend SFFV genomes (34). The identity of restriction enzyme sites between pCI-4 DNA and pCI-3 DNA suggested that the deletion in the former genome may

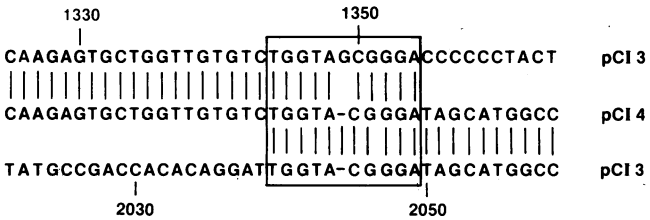


FIG. 6. Location of the deletion within the *env* gene of clone pCI-4. DNA sequences of pCI-3, numbered as in Fig. 3, are compared with those obtained from clone pCI-4. Vertical lines represent homologous nucleotides. The direct repeat is enclosed in a box.

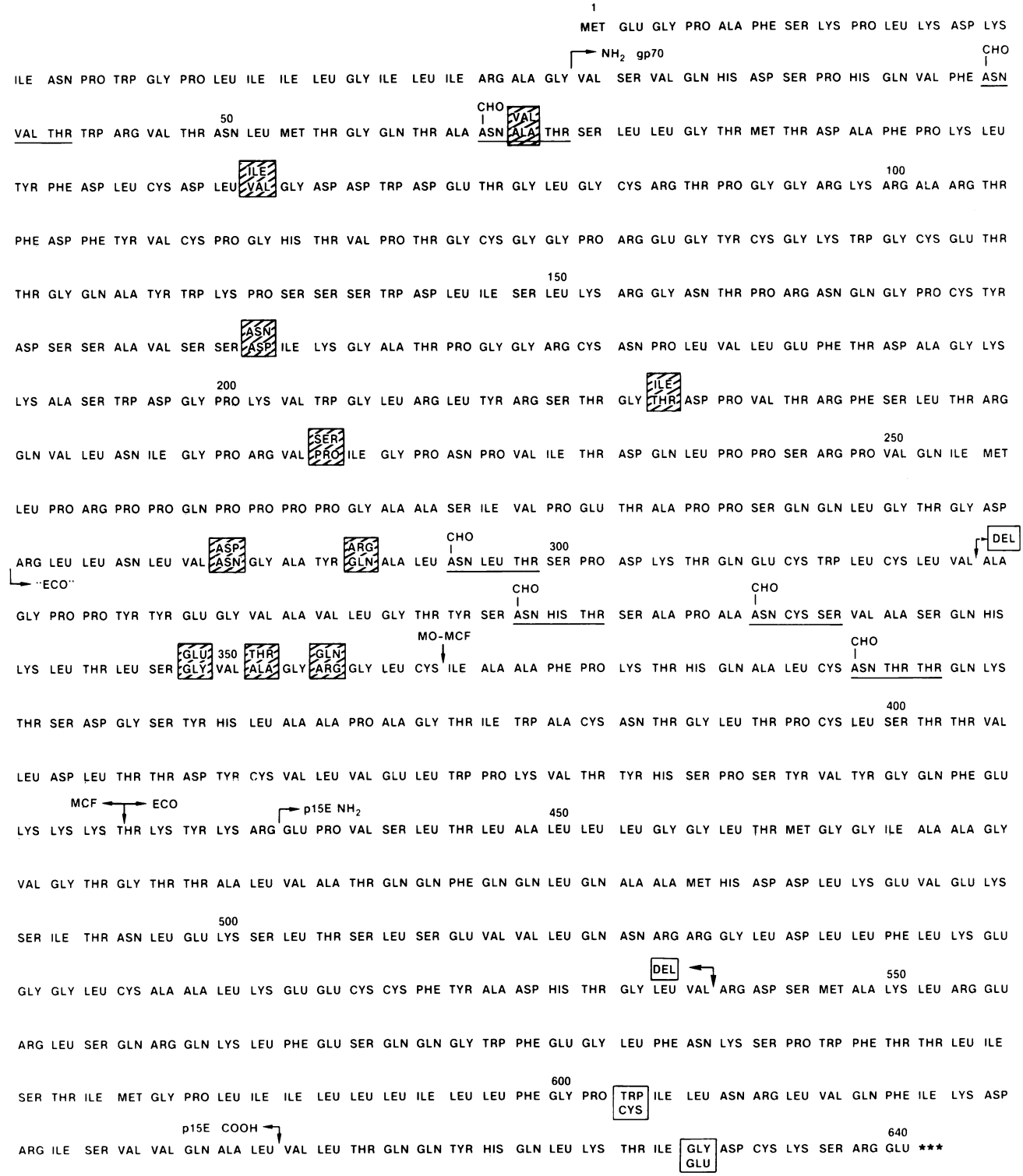


FIG. 7. Comparison of the predicted amino acid sequence of the pCI-3, Mo-MCF, and AKV ecotropic *env* gens. Numbering begins with the methionine initiation codon of the envelope leader sequence. The post-translational cleavage sites which generate the mature gp70 amino terminus and the p15 amino and carboxy termini are indicated with arrows, as are the beginning of the recombinant ecotropic-like sequences ("Eco"), the beginning and end of the sequence deleted in clone pCI-4 (boxed DEL), and the locations of the envelope 3′ recombinant junctions of Mo-MCF (1) and clone pCI-3 shown here (MCF <– –> Eco). Sequence differences within the MCF protion of the pCI-3 *env* gene relative to Mo-MCF are shown in shaded boxes. Differences within the ecotropic p15E sequences relative to AKV are shown in open boxes. Potential glycosylation sites are underlined and contain the abbreviation CHO above.

have been due to an event subsequent to the *env* gene recombination analyzed in the pCI-3 DNA clone. The *Eco*RV site (nucleotides 2,158 to 2,163; Fig. 3) present in the carboxy half of Prp15E was utilized to obtain the nucleotide sequence upstream from this site in pCI-4 DNA. The deletion found in pCI-4, calculated to be 684 nucleotides long, was presumably bordered by a nearly perfect direct repeat (10 of 11 nucleotides) seen in the pCI-3 sequence (Fig. 3). The sequence comparison suggests that the actual crossover event occurred between the repeated TGGTA sequences (Fig. 6). Translation of the pCI-4 *env* gene is not prematurely terminated by the deletion since the reading frame re-entered was that coding for p15E. The mutated protein thus formed is 408 amino acids long, lacking three of the six *env* gene glycosylation sites and the extensive hydrophobic domains which flank the gp70/p15E cleavage site. Since this deletion presumably occurred by homologous recombination between directly repeated sequences, it should be noted that a search of the ecotropic gp70 sequence did not reveal the TGGTA-GCGGGA sequence or other extensive directly repeated regions. We infer, therefore, that a nonecotropic gp70 is a prerequisite of the deletion mutant.

Wolff et al. (39) have recently located the large deletion present in the *env* gene of the polycythemia-inducing form of SFFV. The proposed termini of this deletion are within 60 nucleotides of those described above for pCI-4 DNA. Alignment of the analogous sequences in pCI-3 DNA reveals that smaller, less perfect direct repeats (six of eight nucleotides) flank the coding region deleted from SFFV.

DISCUSSION

We have described here the envelope gene sequences of two in vitro-generated MCF-MuLVs. Comparison on these sequences with those of ecotropic AKR-, Mo-, and Friend MuLV as well as with Mo-MCF revealed several important features. (i) The extent of recombinational exchange in CI-3 is from 145 nucleotides 3′ of the splice acceptor site for the envelope mRNA to nucleotide 1,722, between the end of gp70 and the beginning of Prp15E. No evidence for additional recombinational alteration was obtained. (ii) The *env* gene substitution in CI-3 was virtually identical in sequence (18 changes out of 1,158 nucleotides) to that present in Mo-MCF. The major difference was the location of the 3′ MCF/ecotropic junction (see below). (iii) Comparison of the deduced amino acid sequence of the CI-3 MCF gp70 with that of ecotropic gp70s (either AKV shown in Fig. 5 or Mo and Friend) reveals extensive conservation of the primary sequence (ecotropic-like carboxy half of the molecule) as well as a centrally located polyproline domain which may confer a flexible α-helical conformation to this region (3). Two regions of substantive differences are found on either side of the region of partial homology (see Fig. 5). The CI-3 MCF gp70 sequence contains an additional glycosylation site and an insertion relative to the ecotropic gp70 sequence. Relative to the MCF sequences in CI-3, the ecotropic gp70 contains a polyproline-rich insertion upstream of the partial homology region and an additional glycosylation site downstream. The similarities seen throughout the remainder of the envelope gene suggest that either or both of these two variable regions may encode the receptor specificity and consequently the host range of these viruses. Recently, Repaske et al. (30) reported the sequence of the *env* gene amino terminus from an infectious xenotropic MuLV (NFS-Th-1). The predicted amino acid sequence was 87% homologous to that of CI-3, with the major differences residing in the 25 to 30 amino acids on either side of the region of partial homology. These changes may interfere with the penetration of mouse cells by xenotropic MuLVs. The nonidentity of xenotropic and MCF gp70s also suggests that xenotropic MuLVs are not the nonecotropic parent of the *env* gene of MCF-MuLVs. (iv) The deletion present in the replication-defective CI-4 virus is bordered by an 11-nucleotide direct repeat in CI-3 viral DNA and may therefore be the result of either slippage during reverse transcription or homologous recombination within or between CI-3-like DNA proviruses.

On the basis of restriction endonuclease maps, different AKV-MCF viruses exhibit variation in the extent of the envelope gene substitution (2). In CI-3 the limit of the substitution was located at its 3′ end to 14 nucleotides before the amino terminus of Prp15E. The 5′ border was more difficult to pinpoint and probably was between 100 to 130 nucleotides upstream from the AUG of Prgp85. Our previous measurements of the gp70 substitution in CI-3, as derived from DNA heteroduplexing studies, were shorter at the 3′ end, presumably due to the fact that the carboxy-terminal half of this region contains DNA sequences which are guanine plus cytosine rich (the conserved proline-rich domain) and hybridized to ecotropic AKV viral DNA in spite of only a 53% homology. Thus, it appears that the entire gp70 in CI-3 was derived from an endogenous dual-tropic provirus and the recombinational exchange was initiated and terminated within conserved sequences showing considerable (>90% at the 5′ border) or at least notable (~80% at the 3′ border) homology between endogenous and exogenous envelope genes. The sequence of the Prp15E portion of the *env* gene was probably derived from the ecotropic parent and was not part of the endogenous dual-tropic provirus (only six base changes out of 600 nucleotides were observed). The MCF substitution in Mo-MCF showed a similar location of its 5′ border but did not extend to the carboxy terminus of gp70; instead it terminated at nucleotide 1,480 of the pCI-3 sequence.

Comparison of the substituted sequence in CI-3 with that of Mo-MCF suggests a very close relationship if not identity between the endogenous dual-tropic proviruses from which they were derived. Thus, it appears that one specific provirus out of a large number of endogenous nonecotropic MuLVs is used for the formation of a recombinant MCF-MuLV, independent of whether the recombination occurs in lymphoid tissue in vivo or in fibroblast cells in vitro, or even in different mouse strains (C3H/He versus BALB/Mo). This apparent selectivity may indicate that there is only one very homogeneous family of endogenous MuLV that can provide a dual-tropic host range. Alternately, these endogenous *env* genes may be heterogeneous, and only one can endow the recombinant MCF-MuLV with the cell tropism selected in vivo or in vitro. The cose relatedness of MCF gp70 sequences in different MCF-MuLVs is consistent with the finding that several independent isolates of MCF-MuLV appear to utilize the same cell surface receptor for infection (29).

Mechanism of envelope gene substitution. Two major mechanisms for the formation of MCF-MuLVs may be envisaged; one involes recombination between unintegrated DNA intermediates representing an exogenous viral genome and an endogenous dual-tropic proviral sequence. Alternatively, the recombination event may proceed through the formation of H structures during reverse transcription of heterozygous genomic RNAs (12). We feel that this displacement-assimilation mechanism is consistent with the envelope sequences of the pCI-2 and pCI-3 DNAs (shown here) and their compari-

son to the Mo-MCF envelope sequences (1). Mechanistically, the initial step requires formation, between replicating genomes, of a single-stranded DNA branch (anchored by homologous sequences) as a consequence of plus-strand DNA displacement. This predicts that there are multiple plus-strand initiation sites during reverse transcription, and gaps occur between these sites. The largest heteroduplexed (i.e., recombinant) region would be dictated in part by the distance between plus-strand DNA initiation sites. In accordance with this model, the 5′ recombinant junction of the CI-3 and Mo-MCF genomes represents similar or identical DNA plus-strand initiation sites upon the dual-tropic genome. The 3′ recombinant junction would be within a region of partial homology which preceded a site of DNA plus-strand initiation. There is a paucity of information concerning DNA plus-strand initiation sites. The one site which has been described is preceded by a polypurine sequence (36). The 3′ recombinant junctions of the CI-2 and CI-3 genomes are preceded by purine stretches of 17 and 14 nucleotides, respectively. These junctions, as well as the Mo-MCF/Mo-MuLV 3′ junction, show a clustering of base changes, relative to the CI-3 MCF sequences, which may be the consequence of resolution of closely related but nonidentical sequences following recombinant strand assimilation. Utilizing a procedure which we have shown generates MCF recombinants in vitro (27), we are now in a position to test in vitro some of the predictions of this and other recombination models.

LITERATURE CITED

1. **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.
2. **Chattopadhyay, S. W., 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.
3. **Chen, R.** 1982. Complete amino acid sequence and glycosylation sites of glycoprotein gp71A of Friend murine leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **79**:5788–5792.
4. **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.
5. **Clewell, D. B., and D. R. Helinski.** 1969. Supercoiled circular DNA-protein complexes in Escherichia coli: purification and induced conversion to an open circular DNA focus. Proc. Natl. Acad. Sci. U.S.A. **62**:1159–1166.
6. **Cummings, I. W., J. K. Browne, W. A. Salser, G. V. Tyler, R. L. Synder, J. M. Smolec, and J. Summers.** 1980. Isolation, characterization, and comparison of recombinant DNAs derived from genomes of human hepatitis B virus and woodchuck hepatitis virus. Proc. Natl. Acad. Sci. U.S.A. **77**:1842–1846.
7. **Devare, S. G., U. R. Rapp, G. J. Todaro, and J. R. Stevenson.** 1978. Acquisition of oncogenicity by endogenous mouse type C viruses: effects of variation in *env* and *gag* genes. J. Virol. **28**:457–465.
8. **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.
9. **Famulari, N. G.** 1983. Murine leukemia viruses with recombinant *env* genes: a discussion of their role in leukemogenesis. Curr. Top. Microbiol. Immunol. **103**:75–108.
10. **Famulari, N. G., D. L. Buchhagen, H. D. Klenk, and E. Fleissner.** 1977. Presence of murine leukemia virus envelope proteins gp70 and p15(3) in a common polyprotein of infected cells. J. Virol. **20**:501–508.
11. **Herr, W., and W. Gilbert.** 1983. Somatically acquired recombinant murine leukemia proviruses in thymic leukemias of AKR/J mice. J. Virol. **46**:70–82.
12. **Junghans, R. P., L. R. Boone, and A. M. Skalka.** 1982. Retroviral DNA H structures: displacement-assimilation model of recombination. Cell **30**:53–62.
13. **Kelly, M., C. A. Holland, M. L. Lung, S. K. Chattopadhyay, D. R. Lowy, and N. Hopkins.** 1982. Nucleotide sequence of the 3′ end of MCF 247 murine leukemia virus. J. Virol. **45**:291–298.
14. **Koch, W., G. Hunsmann, and R. Friedrich.** 1983. Nucleotide sequence of the envelope gene of Friend murine leukemia virus. J. Virol. **45**:1–9.
15. **Lee, J. C., and J. N. Ihle.** 1981. Chronic immune stimulation is required for Moloney leukemia virus-induced lymphomas. Nature (London) **289**:407–408.
16. **Lenz, J., R. Crowther, A. Straceski, and W. Haseltine.** 1982. Nucleotide sequence of the AKV *env* gene. J. Virol. **42**:519–529.
17. **Linemeyer, D. L., J. G. Menke, S. K. Ruscetti, L. H. Evans, and E. M. Scolnick.** 1982. Envelope gene sequences which encode the gp52 protein of spleen focus-forming virus are required for the induction of erythroid cell proliferation. J. Virol. **43**:223–233.
18. **Lowy, D. R., E. Rands, S. K. Chattopadhyay, C. 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.
19. **Lung, M. L., C. Hering, J. W. Hartley, W. P. Rowe, and N. Hopkins.** 1980. Analysis of the genomes of mink cell focus-inducing murine type C viruses: a progress report. Cold Spring Harbor Symp. Quant. Biol. **44**:1269–1274.
20. **Marshall, R. D.** 1974. The nature and metabolism of the carbohydrate protein linkages of glycoproteins. Biochem. Soc. Symp. **40**:17–26.
21. **Maxam, A., and W. Gilbert.** 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.
22. **McGrath, M. S., and I. L. Weissman.** 1979. AKR leukemogenesis: Identification and biological significance of thymic lymphoma receptors for AKR retroviruses. Cell **17**:65–76.
23. **Oroszlan, S., and R. Gilden.** 1980. Primary structure analysis of retrovirus proteins, p. 299–343. *In* J. R. Stevenson (ed.), Molecular biology of RNA tumor viruses. Academic Press, Inc., New York.
24. **Pinter, A., and E. Fleissner.** 1977. The presence of disulfide-linked gp70-p15(E) complexes in AKR murine leukemia virus. Virology **83**:417–422.
25. **Queen, C. L., and L. J. Korn.** 1980. Computer analysis of nucleic acids and proteins. Methods Enzymol. **65**:595–609.
26. **Rapp, U. R.** 1980. Isolation of viruses that induce leukemia, sarcoma and carcinoma from mouse fibroblast cell lines chronically infected with endogenous murine leukemia virus, p. 1005–1020. *In* M. Essex, G. Todaro, and H. zur Hausen (ed.), Viruses in naturally occurring cancers. Cold Spring Harbor Conferences on Cell Proliferation, vol. 7. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
27. **Rapp, U. R., E. Birkenmeier, T. I. Bonner, M. A. Gonda, and M. Gunnell.** 1983. Genome structure of mink cell focus-forming murine leukemia virus in epithelial mink lung cells transformed in vitro by iododeoxyuridine-induced C3H/MuLV cells. J. Virol. **45**:740–754.
28. **Rapp, U. R., and T. H. Marshall.** 1980. Cell surface receptors for endogenous mouse type C virus glycoproteins and epidermal growth factor: tissue distribution in vivo and possible participation in specific cell-cell interaction. J. Supramol. Struct. **14**:343–352.
29. **Rein, A.** 1982. Interference grouping of murine leukemia viruses: a distinct receptor for the MCF-recombinant viruses in mouse cells. Virology **120**:251–257.
30. **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.
31. **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.
32. **Rowe, W. P., M. N. Cloyd, and J. W. Hartley.** 1979. Status of the association of mink cell focus-forming viruses with leukemogenesis. Cold Spring Harbor Symp. Quant. Biol. **44**:1265–1268.
33. **Schultz, A., A. Rein, L. Henderson, and S. Oroszlan.** 1983. Biological, chemical, and immunological studies of Rauscher ecotropic and mink cell focus-forming viruses from JLS-V9 cells. J. Virol. **45**:995–1003.
34. **Schultz, A. M., S. K. Ruscetti, E. M. Scolnick, and S. Oroszlan.** 1980. The env-gene of the spleen focus-forming virus lacks expression of p15(E) determinants. Virology **107**:537–542.
35. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
36. **Swanstrom, R., W. J. DeLorbe, J. M. Bishop, and H. E. Varmus.** 1981. Nucleotide sequence of cloned unintegrated avian sarcoma virus DNA: viral DNA contains direct and inverted repeats similar to those in transposable elements. Proc. Natl. Acad. Sci. U.S.A. **78**:124–128.
37. **Temin, H. M., E. Keshet, and S. K. Weller.** 1979. Correlation of transient accumulation of linear unintegrated viral DNA and transient cell killing by avian leukosis and reticuloendotheliosis viruses. Cold Spring Harbor Symp. Quant. Biol. **44**:773–778.
38. **Thomas, C. Y., and J. M. Coffin.** 1982. Genetic alterations of RNA leukemia viruses associated with the development of spontaneous thymic leukemia in AKR/J mice. J. Virol. **43**:416–426.
39. **Wolff, L., E. Scolnick, and S. Ruscetti.** 1983. Envelope gene of the Friend spleen focus-forming virus: deletion and insertions in the 3′ gp70/p15E-encoding region have resulted in unique features in the primary structure of its protein product. Proc. Natl. Acad. Sci. U.S.A. **80**:4718–4722.
40. **Yoshimura, F. K., and K. L. Levine.** 1983. AKR thymic lymphomas involving mink cell focus-inducing murine leukemia viruses have a common region of provirus integration. J. Virol. **45**:576–584.