Truncated gag Products Encoded by Gv-1-Responsive Endogenous Retrovirus Loci†

PAUL F. POLICASTRO,‡ MERETE FREDHOLM,1,2 AND MICHAEL C. WILSON1*

Department of Molecular Biology, Research Institute of Scripps Clinic, La Jolla, California 92037, and Department of Animal Genetics, The Royal Veterinary and Agricultural University, Bülowsvej 13, 1870 Frederiksberg C, Denmark

Received 7 March 1989/Accepted 16 June 1989

The conversion of endogenous or exogenous murine retroviruses to a leukemogenic phenotype involves recombination with retroviral sequences present in host genomic DNA. In the 129 Gis+ inbred strain, these endogenous sequences are replication defective but still express retroviral proteins under the apparent transcriptional control of the Gv-1 regulatory locus. To study the protein-coding potential of Gv-1-regulated endogenous retroviral loci, we used oligonucleotide probes directed to env deletion breakpoints identified in previously characterized cDNA clones. Four endogenous retroviral loci were isolated from a library of 129 Gis+ genomic DNA with these probes. Three loci cloned with the env deletion probe del env-1 had virtually identical proviral inserts by restriction analysis. A unique locus was identified and cloned with the del env-2 probe, which must therefore represent a Gv-1-responsive element. Restriction enzyme and nucleotide sequence analyses indicated that the del env-1 and del env-2 loci represented members of the polytropic and modified polytropic classes of endogenous retrovirus, respectively. Despite this divergence, members of both classes contained identical deletions of 19 nucleotides within p30gen and of 1,474 nucleotides from p10gen into the reverse transcriptase-coding region of pol, suggesting that a recombination event had occurred between these proviral sequences prior to insertion within the genome. The del env-1 and del env-2 loci retained coding capacity for truncated gag polyproteins, confirmed by in vitro translation and immunoprecipitation of the protein products. Nucleotide sequence comparison of the untranslated leader (L) regions of the del env-1 and del env-2 loci to a replication-competent ecotropic virus indicated regions that might be important to dispersion of these endogenous retroviral elements throughout the host genome.

Replication-defective type C retrovirus sequences present in the mouse genome recombine with infectious ecotropic viruses to produce leukemogenic, polytropic recombinant viruses (9, 15, 17, 22, 25). Significantly, these recombinant viruses have exchanged sequence largely including the gp70 coding region of the env gene, which is thought to define retroviral tropism through either attachment and/or uptake of virus particles. The ecotropic viruses infect only mouse cells, while the recombinant polytropic viruses infect cells of both mouse and other species, and xenotropic viruses infect only non-mouse cells. The 30 to 50 endogenous retrovirus loci in any given inbred mouse strain consist of only a few or no ecotropic loci (21) and a variable number (5 to 20 of each) of loci bearing xenotropic (39, 42), polytropic, or modified polytropic env sequences (54). The modified polytropic class contains 18 amino acid changes and a nine-codon deletion within the env coding region relative to the polytropic class, without losing the capacity to contribute functional env sequence to infectious virus (53). The gp70-coding 5' env sequences from a polytropic or modified polytropic locus are thought to contribute one obligate phenotype in the leukemogenic conversion process; the p15E-coding 3' env and adjoining U3 long terminal repeat (LTR) sequences from a xenotropic locus contribute the second (9, 10, 15, 22, 24, 42). Of the many polytropic and modified polytropic loci present in the genome of an inbred strain, not all will be capable of providing the specific env sequences necessary for successful phenotype conversion of a replicating ecotropic virus. In addition to point mutations and short deletions leading to a shift in or loss of the translational reading frame (53), loci and transcripts have been characterized that are deleted for the majority of the env coding region (25, 28, 32). Recombinations involving the env region are easily detected antigenically by changes in the gp70 envelope protein in the leukemogenic conversion process, but endogenous loci deleted for env sequences may act as recombination partners in other regions of the retroviral genome. A Moloney murine leukemia virus (MoMuLV) mutant, normally NB-tropic but expressing a truncated 45-kilodalton (kDa) gag product and containing a 1-kilobase (kb) deletion, was rescued from the JLS-V11-NP cell line as an N-tropic, large XC plaque recombinant, presumably by recombination within gag with the endogenous BALB/c XC-negative, N-tropic retrovirus (5). MoMuLV mutants deleted within either the 3' end of the LTR (12) or the reverse transcriptase-coding region of pol (45) have been isolated as revertants from NIH 3T3 cells bearing new sequences that derive from recombination with one of several endogenous retroviral loci. The frequency of biologically significant recombination between infectious virus and endogenous loci may depend on RNA template exchange during reverse transcription (41) rather than on DNA-mediated recombination involving accessible endogenous loci in host chromatin. Since working models of leukemogenic conversion place env and U3, LTR recombinational events in the spleen and thymus (15, 23, 28), it is possible that endogenous loci involved in the recombination process exhibit tissue-specific expression.

* Corresponding author.
† Publication 5789-MB from the Research Institute of Scripps Clinic.
‡ Present address: Laboratory of Microbial Structure and Function, National Institute of Allergy and Infectious Diseases, Rocky Mountain Laboratories, Hamilton, MT 59840.
We are interested in the regulation of transcription of these endogenous retrovirus sequences and have used the 129 G<sub>va</sub>− and G<sub>va</sub>− strains to study two control loci, Gv-l and Gv-2, that affect endogenous retrovirus expression (50-52). Alternative alleles at these loci behave in a semidominant and dominant fashion, respectively, to regulate endogenous retrovirus transcript (30, 31) and protein (29) levels. The congenic 129 G<sub>va</sub>− strain is homozygous for the alternative b allele at the Gv-l locus and shows 10- to 20-fold lower levels of endogenous retrovirus expression than the G<sub>va</sub>− strain. To study the sequences controlled by Gv-l, we characterized cDNA clones representing transcripts from endogenous retroviral loci in the 129 G<sub>va</sub>− mouse (32). Although the murine strain 129 genome lacks replication-competent ecotropic proviral loci (21), some of the cDNA isolates represent intact polytropic and modified polytropic env genes that could serve in other mouse strains as donors in the ecotropic-to-polytropic conversion process. Other cDNA isolates represent loci deleted for >1 kb of env sequence but are regulated coordinately with those containing full-length env genes. These env deletion breakpoints can thus be used as convenient markers for specific loci that are actively transcribed and regulated by the Gv-l locus, as opposed to the large number of loci that conserve env sequence but are not necessarily transcriptionally active.

In the present report, oligonucleotide probes directed to two env deletion breakpoints were used to isolate provirus-like structural loci from the 129 G<sub>va</sub>− genome. Analysis of these loci by in vitro transcription and translation and by nucleotide sequencing reveals significant deletions of gag and pol domains that result in truncated gag protein products. Three of the isolates contain nearly identical endogenous retrovirus sequence, inserted at different loci in the host genome. In conjunction with the loss of the normal env gene splice acceptor sequence, the gag region encodes the only possible protein products for this family of env-deleted loci. These loci should prove useful in identifying features of host chromatin and endogenous retrovirus structure that result in Gv-l-regulated transcription.

MATERIALS AND METHODS

Oligonucleotide synthesis and labeling. Two nonadecamere oligonucleotides corresponding to the antisense strand of pol-env sequence breakpoints identified (32) in cDNA isolates S13 (5-GAATACAGGGTCTTCTTCT; del env-l) and E2 (5'-TACAAAATGGACCTTTCAT; del env-2) were synthesized by the phosphoramidite method on an Applied Biosystems 380-A DNA synthesizer. (The nucleotide sequence data reported in this paper for del env-l and del env-2 will appear in the EMBL, GenBank, and DDBJ Nucleotide Sequence Databases under accession numbers M26005-121 and M26006-122, respectively.) Crude oligomers were electrophoresed on a 20% acrylamide gel in 44.5 mM Trisborate-1 mM EDTA, pH 8.3 (0.5× TBE), eluted in 4 ml of 0.5 M ammonium acetate–10 mM magnesium acetate, concentrated with n-butanol to 100 μl and precipitated with 1 ml of absolute ethanol. Stocks (5 pmol) were used for end-labeling with [γ−32P]ATP (7,000 Ci/mmol; Amershamp Corp.) by using T4 polynucleotide kinase (New England BioLabs) as described before (35). Labeled oligonucleotides were used directly for library screening and colony hybridization or isolated on 20% acrylamide–8.3 M urea gels for Northern (RNA blot) and Southern hybridization.

Mice, nucleic acid isolation, and Southern and Northern blot hybridization. All mice were obtained at 6 to 8 weeks of age from the mouse breeding colony at Scripps Clinic. DNA was isolated from thymus and liver as described before (7). Genomic DNA was digested with a twofold excess of restriction enzymes under conditions recommended by the manufacturer and analyzed by electrophoresis on 0.7% agarose–2× TAE gels at 1 to 1.5 V/cm, 10 μg of DNA per lane (1× TAE is 40 mM Tris acetate [pH 8.1], 2 mM EDTA). Recombinant DNA digests were separated on 0.9% agarose–0.5× TBE gels. Transfer of DNA to nitrocellulose was performed as described before (47). RNA extraction from tissues, polyadenylated DNA isolation by oligo(dT)-cellulose chromatography, electron microscopy, and sequence determination of env were carried out as described (22). M formaldehyde gels, and transfer to nitrocellulose were performed as described before (30). Prehybridization and hybridization with oligonucleotide probes were performed at 37°C with 6× TESS–50 mM sodium phosphate–5× Denhardt solution (13)–100 μg of denatured salmon sperm DNA per ml–2 mM sodium PP<sub>2</sub>–0.1% sodium dodecyl sulfate (SDS) (20× TESS is 3 M NaCl, 100 mM TES [N-tris(hydroxymethyl)-methyl-2-aminoethanesulfonic acid, pH 7.4], 100 mM EDTA). Labeled oligonucleotides were added to a final concentration of 1.8 pmol/ml at a specific activity of 0.5 × 10<sup>7</sup> to 1 × 10<sup>9</sup> cpm/pmol. After >24 h of hybridization, washes were performed with 6× SSC (1× SSC is 0.15 M NaCl, 0.015 M sodium citrate) or with 3 M tetramethylammonium chloride (59) at temperatures defined in individual experiments below. Prehybridization and hybridization of nick-translated (43) MoMLV poly probe Kpnl fragments (nucleotides [nt] 2858 to 5577) (46) to new or regenerated (56) blots was performed at 42°C in 50% formamide–5× PIPES, 5× Denhardt solution–100 μl of denatured salmon sperm DNA–2 mM sodium PP<sub>2</sub>–0.5% SDS (20× PIPES is 3 M NaCl, 100 mM PIPES [piperazine-N,N'-bis(2-ethanesulfonic acid), pH 8.1], 100 mM EDTA). Washes were performed after >36 h of hybridization.

Library construction and screening and phase DNA isolation. High-molecular weight DNA from 129 G<sub>va</sub>− mouse liver was digested to completion with EcoRI and sedimented on a linear 10 to 40% sucrose density gradient (34), and DNA from 7 to 20 kb in length was pooled after fractions were sized by electrophoresis on 0.7% agarose gels. EcoRI-BamHI-digested EMBL4 DNA (Promega Biotech) was prepared and recommended by the supplier and ligated to the sized mouse genomic DNA EcoRI fragments. Ligated DNA was added directly to bacteriophage lambda DNA packaging mix (Giga-pack; Stratagene), plagues were plated on Escherichia coli NM539, and plagues were transferred to nitrocellulose filters for probe screening (58). Oligonucleotide probes of del env<sub>=</sub> and del env<sub>–</sub>2 were hybridized to library filters under the conditions described above, except that the oligomer concentration was 0.72 pmol/ml. Plagues which cross-hybridized to an endogenous cDNA pol-env nick-translated probe (Psl-EcoRI fragment of E1) (32), under conditions described for Southern and Northern blot analysis, were plaque purified with the del env probes. DNA was extracted from phage isolated by DE52 chromatography of broth lysates (19). EcoRI inserts were subcloned in pUC19 for restriction enzyme mapping and hybridization after preliminary characterization in the EMBL4 vector. Subclones were probed with fragments of the del env-2 cDNA E2 (32) representing 3′ pol-p15E (HindIII-PstI) and U<sub>3</sub> LTR (PstI-Kpnl) sequences.

In vitro transcription and translation and analysis of protein products. To generate capped RNA transcripts, Kpnl fragments extending from the 5′ LTR region to the 3′ LTR and encompassing the entire endogenous retroviral genomic loci

Vol. 63, 1989

Gv-l-RESPONSIVE ENDOGENOUS RETROVIRAL LOCI 4137
3.1, 3.2, 11.1, and 15.3 were subcloned into the pSP18 expression vector (Bethesda Research Laboratories). Similarly, KpnI fragments of MCF 247 and MX27 (generously provided by Christine Holland and Jonathan Stoye, respectively) extending from the conserved KpnI site in the 5' LTR to the KpnI site within the 5' region of pol were also subcloned in pSP18. The plasmids were linearized with either HindIII or EcoRI distal to the SP6 RNA polymerase initiation site, depending on the orientation of the insert in the plasmid vector. Transcription reactions were performed with SP6 as described before (16) in the presence of 500 µM 7mGpppG" (Pharmacia), 100 mM GTP, CTP, ATP, and UTP, and 5 µCi of [α-32P]GTP (3,000 Ci/mmol; New England Nuclear Corp.). The transcripts were extracted with phenol-chloroform, passed over G-50 Sephadex, and ethanol precipitated before use in a translation assay consisting of 18 µl of rabbit reticulocyte lysate (Amersham) and 20 µCi of [35S]methionine (1,127 Ci/mmol; New England Nuclear) programmed by 200 ng of RNA transcript. The translation reactions were stopped by adding equal volumes of 0.1 M NaCl, 20 mM Tris (pH 7.0) and 1% Triton X-100, and 2.5 mM phenylmethylsulfonyl fluoride. To assay the translation product directly, 3 µl of the reaction mixture containing stop solution was diluted with 45 µl of sample buffer (2.5% SDS, 5% β-mercaptoethanol, 20% glycerol, 0.125 M Tris (pH 6.8)), boiled for 10 min, and applied to 10% SDS–polyacrylamide gels (27). For immunoprecipitation of the in vitro-synthesized protein products, 3 to 5 µl of goat antiserum raised against Rauscher MuLV p15, p12, and p30 or normal goat serum (provided by John Elder) per sample was preabsorbed to protein A-Sepharose (Pharmacia) in PORT buffer (150 mM NaCl, 10 mM Tris [pH 8.5], 100 µg of ovalbumin per ml, 0.1% Nonidet P-40, and 0.5% Tween-20) for 2 h at 4°C and washed in PORT buffer before incubation with the reaction mixture at 4°C overnight. The protein A-Sepharose-antibody complex was washed with PT buffer (150 mM NaCl, 10 mM Tris [pH 8.5]) before elution by boiling in sample buffer. The translation products after electrophoresis were visualized by impregnating the gel with Enhance (New England Nuclear) and autoradiography.

**Nucleotide sequence analysis.** The dideoxy nucleotide chain termination method (44) with [3H]dideoxyadenosine 5'-(α-thio)-triphosphate used for detection was applied to pUC19 subclones of the EMBL4 inserts prepared by alkaline denaturation (3) or by further subcloning into single-stranded bacteriophage M13mp10 and mp11 (35); standard M13 primers and custom primers defined by internal sequences, synthesized as described above for the env deletion probes, were used to initiate reactions on the templates. Reaction products were resolved on 6% acrylamide buffer gradient gels (6).

**RESULTS**

**env-deleted endogenous retroviral transcripts and proviral loci.** We have previously described the structure and sequences of cDNAs encoding the env and U3 LTR regions of endogenous retroviruses as isolated from the spleen, liver, and epididymus of 129 G_{1}x_{1} inbred mice (31, 32). A number of independent cDNA isolates displayed extensive deletions of env that we have categorized as either del env-1 or del env-2 sequences. The S13 cDNA contains the del env-1 deletion, extending from nt 5676 at the 3' end of pol to nt 7673 in the env p15E-coding region, while the E2 cDNA represents the del env-2 deletion, extending from nt 5796 at the 3' end of pol to nt 7696 in the env p15E sequence (corresponding to AKV virus nucleotide sequence nomenclature [20]). Each of these deletions was apparently mediated by a mechanism that involves recognition of direct heptamer nucleotide repeats present in full-length endogenous retrovirus sequences at the deletion breakpoints. We designed oligonucleotide probes to the novel sequence juxtaposition created by these deletions, each with a length (19 nt) that would not permit competitive hybridization to either end of the sites in nondeleted retroviral sequences. In this way the oligonucleotide probes discriminate between loci bearing these deletions and the numerous endogenous retrovirus loci present in the 129 G_{1}x_{1} mouse genome. The selected loci would thus include those particular endogenous retrovirus sequences, responsive to Gv-L control, from which transcripts arose. We first tested the specificity of the oligonucleotides on endogenous retrovirus transcripts from 129 G_{1}x_{1} tissues and compared the hybridizing species with the RNAs identified with an M0MuLV p0l region probe (Fig. 1). The del env-1 probe identified a 5.0-kb transcript in thymus and a 4.0-kb transcript in liver (Fig. 1A), while del env-2 identified a 5.2-kb spleen transcript (Fig. 1B). These transcripts were described previously, by size and tissue of origin, with the use of a full-length M0MuLV clone (30) and a 129 G_{1}x_{1} endogenous retroviral pol-env cDNA clone (31) as probes; levels of these transcripts were shown to be reduced 10- to 20-fold in 129 G_{1}x_{1} mouse tissues, by Northern blot hybridization analysis and with S1 nuclease protection assay. In addition to these major del env-1 and del env-2-hybridizing species, transcripts of similar size were present in lesser abundance in each tissue tested. In contrast, rehybridization of the del env-2 blot to the M0MuLV p0l region probe showed that retrovirus-related RNA species were present in sizes from 8.0 to 3.5 kb in these three tissues (Fig. 1C).
showed that del env-1 transcripts could come from any of four different loci, on 13.5-, 10.0-, 9.0-, and 8.8-kb EcoRI fragments (Fig. 2B); a fifth hybridizing fragment, at 8 kb, was not retained at slightly higher wash stringency (data not shown). However, del env-2 transcripts appeared to have only one source, present on an 8.9-kb EcoRI fragment (Fig. 2C). Identical hybridization patterns by the 129 Gs\textsuperscript{+} and congenic 129 Gs\textsuperscript{−} DNAs (Fig. 2B and C) demonstrated that specific proviral loci responsible for Gv-1-regulated transcripts were retained in the 129 Gs\textsuperscript{−} congenic strain, which carries the Gv-1\textsuperscript{+} allele of the nonpermissive C57BL/6 strain (51).

**env-deleted loci contain a common gag-pol deletion.** To isolate the loci identified by the del env oligonucleotides, the 7- to 20-kb size-fractionated EcoRI fragments of 129 Gs\textsuperscript{+} DNA were ligated into the EMBL4 bacteriophage vector, and 4.2 \times 10\textsuperscript{5} plaques of the resulting library were screened with pooled del env-1 and del env-2 oligonucleotide probes. Phage DNAs cross-hybridizing with the 129 Gs\textsuperscript{+} endogenous pol-env region cDNA clone were mapped with restriction enzymes and analyzed on Southern blots for hybridization with separate del env-1 and del env-2 oligonucleotide probes and with U3 LTR and pol-p1SE probes from the del env cDNA clone E2 (see Materials and Methods). Three del env-1 loci, 3.1, 3.2, and 11.1, were present on the 13.5-, 10.0-, and 9.0-kb fragments, respectively, identified in genomic Southern analysis (Fig. 3A). The 3.1 locus was represented by five independent isolates, 3.2 by one isolate, and 11.1 by three isolates. None of the del env-1 clones recovered represented the 8.8-kb genomic EcoRI fragment identified by the oligonucleotide probe (Fig. 2B), although a fragment of this size would be compatible with the size requirement of the EMBL4 phage vector. The single del env-2 locus, 15.3, was represented by three distinct isolates that each gave the same 8.9-kb EcoRI insert and internal restriction map (Fig. 3A). The three del env-1 loci were essentially identical by restriction enzyme analysis within the proviral domains, yet flanking host DNA showed no similarities by the same criteria. By comparison with published endogenous type C murine retrovirus maps (25, 53), (Fig. 3B), the del env-1 and del env-2-2 loci codetermined approximately 1.4 kb of sequence in addition to the env deletions. The distance between conserved U3 PstI sites was 4.8 and 5.0 kb for the del env-1 and del env-2 clones, respectively, but sizes of 6.3 and 6.4 kb are to be expected if the proviruses lack only the env sequences. Although the transcript size from these loci appeared to be approximately 5 kb for transcripts arising in spleen and thymus, a 4-kb del env-2 RNA was found in liver (Fig. 1). This may be transcribed from the uncloned del env-1 locus on the 8.8-kb EcoRI genomic DNA fragment or from any one of the del env-1 loci expressed preferentially in the liver and containing novel transcript-processing signals. To resolve the discrepancy between expected and actual proviral insert size, EcoRI inserts bearing the three del env-1 loci and one del env-2 locus were subcloned in pUC19 for further mapping and sequence analysis. Deletion of a gag-pol junction region was suggested by the conservation of 5' gag Xbal and SacI sites and a 3' pol XbaI site in subclones of all four isolates (Fig. 3) and by loss of SacI, KpnI, and BamHI sites expected in the 5' end of the full-length endogenous pol region (25, 53).

To test the gag-pol junction region for the presence of deletions, del env-1 clone 3.1 and del env-2 clone 15.3 were sequenced from a conserved SphI site in the pol domain. The nucleotide sequence revealed identical deletions of gag-pol sequence between nt 2188 and 3675, relative to AKV virus

---

**FIG. 2.** Genomic DNA blot analysis with oligonucleotide probes specific to env-deleted endogenous retrovirus loci. EcoRI-digested high-molecular-weight DNA separated on 0.7% agarose gels was transferred to nitrocellulose and hybridized to 32P-labeled DNA probes. (A) Autoradiograph of C57BL/6 (C57, lane 1) and 129 Gs\textsuperscript{+} (129, lane 2) thymus DNA hybridized to nick-translated MoMuLV (Mo) pol fragment, washed at 42°C in 0.1 x SSC-0.1% SDS, and exposed for 20 h on preflashed AR film (Eastman Kodak). \(\phi X174\) HindIII-digested bacteriophage lambda c1857 molecular weight markers. (B) Autoradiograph of 129 Gs\textsuperscript{+} (+, lanes 4 and 5) and 129 Gs\textsuperscript{−} (−, lane 6) liver DNA hybridized to end-labeled 19-mer oligonucleotide del env-1 washed at 54°C in 3 M tetramethylammonium chloride-50 mM Tris (pH 8.0)-2 mM EDTA-0.1% SDS and exposed for 6 days on preflashed XRP film. (C) Autoradiograph of 129 Gs\textsuperscript{+} (+, lane 7) and 129 Gs\textsuperscript{−} (−, lane 8) thymus and 129 Gs\textsuperscript{−} (−, lane 9) liver DNA hybridized to end-labeled oligonucleotide probe del env-2, washed, and exposed as in panel B. Arrows indicate major hybridizing EcoRI fragments present in genomic DNAs.
genome coordinates (20), in both endogenous isolates (data not shown). This 1,474-nt deletion, named del pol, extended from the 3′ end of gag p10 into the reverse transcriptase-coding region of pol (Fig. 3). As with the del env-1 and del env-2 deletions (32) and retroviral and cellular deletions observed by others (14; summarized in reference 28), the del pol junction retained one copy of a 7-bp direct repeat found at the deletion breakpoints in the full-length AKV virus sequence.

Other distinguishing features of the 15.3 restriction map in comparison to those of 3.1, 3.2, and 11.1 were the 3′ pol HindIII site and the 5′ env BamHI site in the del env-2 clone 15.3 and the relative size of the U3 LTR regions. The presence of the 15.3 HindIII site indicated that this insert was derived from a full-length provirus of the modified polytropic class of endogenous retrovirus, while the absence of the pol HindIII sites from the del env-1 clones 3.1, 3.2, and 11.1 distinguished them as polytropic-related sequences (53, 54). The 15.3 BamHI site within env is conserved in endogenous, nonxenotropic retroviruses as a group and was presumably also present in the nondeleted precursors to the del env-1 loci. Double digestion of the genomic isolates with PstI and KpnI showed that the U3 LTR regions of del env-1 clones were ~50 nt shorter than the del env-2 U3 regions (550- versus 600-nt PstI-KpnI fragments; data not shown). The U3 size differences are presumably due to deletions in the single-copy enhancer homolog in del env-1 clones relative to the del env-2 clone, similar to those reported previously in cDNA sequence analysis (31, 33). Therefore, these isolates share identical gag-pol deletions yet are significantly divergent in LTR and env sequences, suggesting independent but equivalent deletion events in the two, or an ongoing genetic exchange as defective proviruses.

**Truncation of gag ORF yields products of 32 and 34 kDa.** One consequence of the 1,474-nt gag-pol deletion in the del env-1 and del env-2 clones is the truncation of the gag open reading frame (ORF) and possibly the production of a novel gag-pol fusion product. The sequence juxtaposition arising from the deletion predicts an ORF extension of 10 codons beyond the breakpoint before a new termination site corre-
sponding to nt 3701 to 3703 of the AKV virus sequence. To test these defective loci for their gag region coding potential, KpnI fragments of 4.8 (del env-1) and 4.9 (del env-2) kb, extending from the 5' U, LTR to the equivalent restriction site in the 3' LTR, were subcloned in a vector which allows SP6 RNA polymerase transcription of inserted sequences. Transcripts from each reaction were capped with 7mGppp in vitro and translated in a rabbit reticulocyte lysate in the presence of radiolabeled methionine. Transcripts from 3.1, 3.2, and 11.1 directed translation of a 34-kDa protein, and transcripts from 15.3 directed translation of a 32-kDa protein (Fig. 4A). If these proteins originated at the equivalent ATG utilized by AKV virus, their apparent molecular mass on SDS-polyacrylamide gel electrophoresis suggests products that extend halfway through the p30 region of gag, since they are half the size expected if terminated in pol as described above. The translation products from 3.1 and 15.3 were immunoprecipitated with individual p15, p12, and p30 Rauscher gag protein antisera to determine whether the 34- and 32-kDa products indeed represented the expected coding region (Fig. 4B). These antisera showed that the 34- and 32-kDa products contained p15 and p12 determinants (32-kDa protein reactivity with p12 antisera was observed after a longer exposure of the autoradiogram shown in Fig. 4B). In contrast, only the del env-1 locus 3.1 product conserved minor antigenic reactivity with the Rauscher p30 antisera. An MCF 247 gag p65 translation product demonstrated uniform recovery with each antisera in parallel. Thus, the del env-1 and del env-2 transcripts encode similar, but not identical, truncated gag protein products that are less than 280 residues long, whereas the gag-pol deletion they share predicts products ~520 amino acids in length. Virus-negative NIH Swiss mice express p12 and p30 liver antigens that cross-react with anti-AKR MuLV and anti-Rauscher MuLV sera but are distinct from the homologous viral antigens (48, 49). BALB/c and 129/J tissues contain p30 but not p12 reactivity to anti-AKR virus sera (36). Based on the abundance of transcripts from the del env-1 and del env-2 loci in 129 G,` mouse tissues (Fig. 1) (30) and the ability of SP6-derived transcripts of the cloned loci to direct translation of products reactive with Rauscher gag protein antisera, we suspect that the del env-1 and del env-2 loci are responsible for a portion of the gag antigen observed in 129 G,` tissues.

We were interested in the possibility that even full-length endogenous proviruses might contain similar cryptic gag region termination signals. To test this, we compared the 3.1 and MCF 247 cell-free products with the product from MX27, a polytropic provirus clone from HRS/J mice that is apparently intact, by restriction enzyme mapping (53). This comparison (Fig. 4C) showed that MX27 encoded a polypeptide of 63 kDa, only slightly smaller than the product from the infectious MCF 247 clone. Although this suggests that the MX27 gag region does contain a premature termination signal or an in-frame deletion relative to MCF 247, it also demonstrates that gag products of 32 to 34 kDa are not a common feature of all endogenous polytropic and modified polytropic proviruses.

Nucleotide sequence confirms structure of gag polypeptides. To identify the ORFs resulting in the observed gag protein products from the del env-1 and del env-2 clones, we determined 5' gag nucleotide sequences for 3.1 and 15.3. Since the 5' leader (L) region of gag has been reported for only one other endogenous murine polytropic provirus, AL10 (40), the L regions were included in the 3.1 and 15.3 analyses. Figure 5 shows these data, beginning with the R region of the 5' LTR and extending 1,435 nt through the apparent ORFs of the del env-1 and del env-2 gag regions. The previously reported partial AL10 sequence is included for comparison of L regions (see below). Analysis of the gag protein-coding sequence showed that, relative to the AKV ecotropic virus genomic sequence (20), 3.1 and 15.3 shared
TABLE 1. U3-gag sequence conservation between del env loci and AKV virusa

<table>
<thead>
<tr>
<th>Region\b</th>
<th>Nucleotide positions</th>
<th>Comparison (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>3.1-AKV Div</td>
<td>3.1-AKV Cons</td>
</tr>
<tr>
<td>R</td>
<td>1-69 7.2 92.8 4.3 95.7 4.3 95.7</td>
<td></td>
</tr>
<tr>
<td>U\b</td>
<td>70-146 10.4 89.6 6.4 93.6 6.5 93.5</td>
<td></td>
</tr>
<tr>
<td>PBS</td>
<td>145-194 33.3 60.6 27.8 72.2 20.2 80.1</td>
<td></td>
</tr>
<tr>
<td>5' L</td>
<td>165-240 12.6 87.3 13.8 86.1 1.8 98.3</td>
<td></td>
</tr>
<tr>
<td>3' L</td>
<td>421-667 42.1 57.8 34.9 65.1 7.1 92.7</td>
<td></td>
</tr>
<tr>
<td>p15</td>
<td>688-1054 15.5 84.5 18.9 81.1 9.3 99.0</td>
<td></td>
</tr>
<tr>
<td>p12</td>
<td>1055-1309 22.7 77.3 22.7 77.7 2.3 97.7</td>
<td></td>
</tr>
<tr>
<td>p30</td>
<td>1310-1435 26.1 73.9 27.0 73.2 0.8 99.2</td>
<td></td>
</tr>
</tbody>
</table>

\a Comparisons of del env-1 clone 3.1 and del env-2 cloned 15.3 sequences with those for comparable stretches of the AKV ecotropic virus genome; nucleotide numbering is as shown in Fig. 5.

\b R, Redundant region; U, unique 3'-flanking region in 5' LTR; L, untranslated leader region of gag; p15, p12, and p30, protein-coding regions of gag.

d) Divergence (Div) calculated as number of gaps introduced to create alignment plus number of nonidentical nucleotides divided by total nucleotide length, expressed as percent. Conservation (Cons) calculated as number of identical nucleotides divided by total nucleotide length, expressed as percent.

identical 19-nt deletions in the 5' end of the p30-coding region, between nt 1369 and nt 1389 of the aligned sequences in Fig. 5 and corresponding to nucleotides 1341 through 1359 of AKV virus. This deletion occurred 20 codons into the p30 coding region, and the frameshift resulted in a 10-codon extension of the 3.1 reading frame before termination. The 15.3 clone contained a single-base-pair deletion upstream within the p12-coding sequence at nt 1236 (Fig. 5); the resulting frameshift encoded 44 amino acids in an alternate ORF that terminated at the 19-nt deletion within p30. Besides single-nucleotide substitutions relative to the AKV virus sequence, there were also a 21-nt deletion between nt 798 and nt 820 in the 15.3 ORF and a 3-nt deletion at nt 1179 to 1183 in both the 3.1 and 15.3 sequences. These had a neutral effect on the gag ORF but served to reduce the relative size of the 15.3 product.

These sequence analyses explain both the mass differences of the 3.1 and 15.3 gag products (Fig. 4A) and the Rauscher anti-gag protein antisemur reactivities of these products (Fig. 4B). The 3.1 and 15.3 products should have respective masses of approximately 29.8 and 27.2 kDa, by the derived amino acid sequence shown in Fig. 5. As represented by clone 3.1, the del env-1 family conserved gag protein-coding capacity relative to 15.3 by virtue of the single-nucleotide deletion of the del env-2 clone at nt 1236 in the p12-coding region. Table 1 summarizes the 3.1 and 15.3 polytropic and AKV virus ecotropic sequence comparisons for gag polypeptide-coding regions and for upstream untranslated regions shown in Fig. 5. On the nucleotide level, 15.3 did not show greater gag region divergence from AKV virus than did 3.1; both polytropic clones conserved greater than 80% of the p15-coding sequence relative to AKV virus and equivalent loss of conservation (77 and 73%, respectively) through the p12- and p30-coding sequences. In the same regions, 3.1 and 15.3 maintained greater than 90% sequence identity to each other. This relatedness is consistent with other data on ecotropic-polytropic divergence (54), and emphasizes the significance of the deletion at nt 1236 to coding potential in the del env-2 clone 15.3.

PBS and 3' leader region are not conserved. The R and U\b regions of the 3.1 and 15.3 5' LTRs showed relatively strong (~90%) conservation of sequence compared with AKV virus (Table 1), whereas the untranslated leader (L) sequence within gag showed greater than 20% divergence compared with AKV virus. This divergence was seen first in the presumed tRNA primer-binding site (PBS) for initiation of minus-strand reverse transcription, nt 147 to 164 in Fig. 5, in which 3.1 showed 6 bp changes and 15.3 showed 5 bp changes relative to AKV virus. In murine type C retroviruses in general, the PBS conserves complementarity for the donor tRNA (12). However, the PBS sequence of the present 15.3 clone, of a replication-competent revertant of an MoMuLV deletion mutant recovered from NIH 3T3 cells (12), and of the AL10 endogenous polytropic clone reported previously (40) and shown in Fig. 5 all contained a 17 bp match to the 3' terminus of a rat glutamine tRNA. Clone 3.1 diverged from complementarity to this tRNA primer at nt 18 nt, in addition to its divergence from AKV virus sequences, suggesting that it cannot bind to either primer.

We examined the PBS sequences of del env-1 clones 3.2 and 11.1 to determine whether these clones share this divergence sequence with 3.1. The results of this comparison are shown in Fig. 6. The 3.2 and 11.1 isolates showed perfect complementarity to the glutamine tRNA and were identical to the PBS of an endogenous retrovirus isolate from an RFM/Un strain mouse (38). Divergence of the 3.1 PBS may indicate that this locus is no longer capable of helper virus-dependent replication, whereas the two other del env-1 loci and the del env-2 locus may maintain this capacity. Since we suspected that these PBS modifications in del env-1 clone 3.1 were
errors by helper virus reverse transcriptase, we examined the host-proviral junction sequences for each of the cloned loci to determine whether integration at the 3.1 locus was possibly defective due to helper function. The sequences (Fig. 7) indicated that each insertion event resulted in duplication of 4 to 5 bp of host sequence, as expected from normal integrase activity (56).

Three additional functions encoded by the L region are the splice donor sequence required for env mRNA processing and the dimerization and encapsidation sequences required for genome incorporation into the virion (11). The splice donor sequence identified in AKV virus (20), at nt 205 to 211 (Fig. 5), was fully conserved in the 3.1 and 15.3 clones. Data from 129 Gv-" strain cDNA clones indicate that the splice acceptor upstream of env is part of the 1,997-base-pair (bp) pol-env sequence deleted in del env-1 isolates, while this acceptor is intact and fully conserved upstream of the 1,990-bp deletion in del env-2 clones (32). These transcript-processing signals may be recognized in vivo, as suggested by the 1.7-kb del env-2-specific transcript present in spleens of 129 Gv-" mice (Fig. 2B).

The splice donor sequence was contained in a part of the L region that extends from the 3' end of the PBS at nt 165 to nt 420 (Fig. 5), which is generally conserved (787%, Table 1) between the endogenous isolates and AKV virus. Dimerization of full-length retroviral transcripts involves joining at the 5' ends of each molecule, which may involve this 5' portion of L, including the PBS (11). Efficient encapsidation also requires conservation of sequences within L, although additional sites within gag may enhance recognition (1, 2, 4). These processes may allow packaging of transcripts from the del env-1 and del env-2 loci in the presence of helper virus if L region sequences 5' of nt 420 are the major determinants. Sequence conservation was significantly altered from nt 421 to nt 667, however, with an overall sequence identity of 60% between either of the endogenous isolates and AKV virus (Table 1). Comparison of the previously published endogenous AL10 sequence with that of AKV virus in the 5' end of the L region revealed 85.5% conservation of the sequence from nt 165 to 420, approximately equal to that shown by the 3.1 and 15.3 loci. However, in the 3' segment of L, AL10 contained a 68-nl deletion and a 24-nl insertion which contributed to conservation of only 38.5% of the sequence between nt 421 and 667. Most of the conserved sequence in the latter comparison lay between nt 498 and 596, with strongly conserved sequence from nt 535 to 579. This was also the most conserved stretch in comparisons between AKV virus and the present del env-1 and del env-2 clones, suggesting that this portion of the L region is involved in dimerization or encapsidation.

**DISCUSSION**

We anticipated that design of oligonucleotide probes for two cDNA-encoded endogenous retroviral env region deletions would result in isolation of two distinct, Gv-L-regulated proviral loci from the 129 Gv-" host genome. In fact, the del env-1 probe identified a family of four loci, at least three of which contained essentially the same endogenous polytropic proviral sequences. In the del env-1 family, at least, conservation of the LTR promoter and L region sequences seems to be adequate for helper virus-dependent replication and reinsertion in the host germ line. One recent report (18) suggests that defective retroviruses can in fact undergo intracellular transposition at measurable frequencies. The multiplicity of del env-1 loci causes confusion, however, with regard to the source of Gv-L-regulated transcripts. Since different-sized del env-1 RNAs were observed in thymus and liver, it seems likely that there are internal differences between these loci, in addition to the PBS alterations detected between the 3.1 locus and the 3.2 and 11.1 loci. It also seems likely that tissue-specific factors are responsible for differential expression from the loci generating 5-kb and 4-kb transcripts; whether these factors represent elements in cis, such as chromatin configuration at the proviral insertion, or in trans, such as diffusible DNA-binding proteins, is a matter for further investigation. In the case of the sole del env-2 locus in the 129 genome, there can be only one source for the 5.2-kb transcript observed in the spleen, and it therefore seems likely that this particular insertion site and the U3 LTR sequence of 15.3 are both permissive to expression in a certain population of spleen cells. It should be noted that the 3' U3 LTRs from both del env-1 and del env-2 DNA functions as promoters in an embryonal carcinoma cell line (33). The transcription patterns seen in Fig. 1 only identify relative levels of in vivo expression, since the del env-1 deletion has been found in spleen and liver cDNAs and the del env-2 deletion has been found in epididymus and liver cDNAs (31, 32; unpublished observations).

With the identification of 15.3 as the only possible source of del env-2 transcripts, we were able to corroborate other lines of evidence indicating that Gv-L regulation acts in trans. Although retroviral probes hybridized to RNA from 129 Gv-" tissues detect the same array of transcripts as seen in 129 Gv-" tissues, signal intensities are 10-fold lower or less (30). It is possible that a few endogenous loci responsible for most transcripts in 129 Gv-" mice were deleted in the process of selecting the Gv-L phenotype seen in the 129 Gv-" congenic mouse. However, the del env-2 probe recognized the same 8.9-kb EcoRI fragment in 129 Gv-" DNA as in 129 Gv-" DNA, showing that a transcriptionally active, Gv-L-regulated endogenous retroviral locus was not excised or lost during the derivation of the congenic strain carrying the alternative Gv-L regulatable allele (31).

The polytropic and modified polytropic endogenous viruses appear to have diverged most recently of the four recognized classes of type C murine retroviruses, based on comparison of env gene-coding potential (53). Besides the expected similarities in restriction site maps between the del env-1 and del env-2 loci, we found that they shared identical gag-pol deletions mediated by direct heptamer repeats at the deletion breakpoints and identical gag p30 region 19-nt deletions which interrupted the gag ORF. Since these vi-
ruses are members of the polytropic and modified polytropic classes, one possibility is that these deletion events occurred independently. If so, it is also possible that the del env-l loci, appearing to originate from a single, multiply-deleted provirus, could each have arisen independently from different polytropic viruses undergoing equivalent deletion events during replication. Alternatively, the shared gag-pol deletion could have arisen from RNA template exchange between a transcript of a polytropic gag-pol-deleted, 3.1-like locus and a transcript from a progenitor of the 15.3 polytropic locus bearing only the del env-l deletion, during minus-strand reverse transcription. If minus-strand DNA transcription terminated prematurely within the pol gene of the 15.3 template and was completed on a 3.1 RNA template, plus-strand DNA synthesis would then generate a molecule with the U₃ LTR sequence (41). In this model, the diversification of the 3.1-like loci by retrovirus-mediated transposition and accumulation of additional mutations by the integrated recombinant 15.3 locus would then occur independently.

The analysis of the coding potential of these loci was facilitated by the use of in vitro transcription and translation, followed by immunological identification of products. The retroviral gag gene is particularly amenable to this method because transcript splicing is not required to allow recognition of the initiator AUG and complete ORF translation by the reticulocyte system. Also, the high degree of sequence conservation in this region of the type C retroviral genome permits the use of antibodies raised against Rauscher virus, encoding polypeptides of presumed ecotropic origin, to determine conservation of antigenic determinants. Loss of the env gene splice acceptor site in the del env-l family prevents the dominant mode of retroviral transcript processing, ensuring that the truncated 34-kDa molecule we describe is the major protein product from these loci. In the case of the del env-2 locus 15.3, splice donor and acceptor sites are conserved and apparently functional. This leads to significant levels of both the 5.2-kb unspliced transcript directing translation of the 32-kDa protein and the 1.7-kb transcript that probably contains only 200 nt of p15E-coding sequence from the 5' end of the env ORF. Preliminary evidence suggests that 129 Gₛₛ mice do contain gag products similar to those translated in vitro. Immunoblots of protein extracted from 129 Gₛₛ tissues reacted with anti-Rauscher p12 serum to show the presence of a 30-kDa antigen in spleen and a 33-kDa antigen in thymus (M. C. Wilson, unpublished), consistent with the size and tissue-specific expression of the respective cell-free product sizes (Fig. 5) and tissue-specific expression (Fig. 1) of the del env-2 and del env-l loci. We have not yet identified the coding sequence changes which result in the apparent truncation of the MX27 gag product (Fig. 4) relative to the MCF 247 product. The size difference of ~4 kDa (63 versus 67 kDa) suggests that the p10 ORF is interrupted, but immunological characterization will be necessary to determine whether p15, p12, and p30 determinants are present, since an insertion or deletion mutation could result in a long extension in another ORF.

There is little prior evidence of gag sequence variation and its effect on infectious retrovirus functions, and the isolates characterized here may be useful in creating recombinants with infectious ecotropic virus. It has been demonstrated that an endogenous proviral sequence bearing a PBS compatible with tRNA₈⁸ can serve as a PBS donor in rescue of a defective MoMuLV genome (12). The 3 of 18 nt base mismatch of the 3.1 locus PBS with tRNA₈⁸ does not rule out the possibility that it is compatible with another unidentified primer. Recombinants containing the extensive 3' L region or more conservative p15 and p12 gag sequence differences reported here might have significantly altered phenotypes as infectious virus. The expression of these endogenous sequences and the presence of defective gag gene products do not seem to block replication by infectious retrovirus (57). Conversely, the stretches of conserved sequence in the L region noted above may be adequate to ensure encapsidation of endogenous defective retroviral transcripts in the presence of replication-competent genomes.

ACKNOWLEDGMENTS

We thank J. H. Elder for valued discussions and for critical comments on the manuscript, as well as the Rauscher antisera used in this study. We also thank Jonathan Stoye and John Coffin for the MX27 and Christine Hollare for the MCF 247 provirus containing clones, and we thank Michelle Dietrich for expert secretarial assistance.

This work was supported in part by Public Health Service grants CA-33730 to N.C.W. from the National Institutes of Health, post-doctoral fellowship CA07788 to P.F.P., and a Danish National Research Council (SJVF) grant to M.F.

LITERATURE CITED

evidence that the original Rous sarcoma virus was replication defective. J. Virol. 55:728–735.


