Characterization of a Molecular Clone of RFM/Un Mouse Chromosomal DNA that Contains a Full-length Endogenous Murine Leukaemia Virus-related Proviral Genome

By KAVE N. NIKBAKHT,1 LAWRENCE R. BOONE,2,3 PAUL L. GLOVER,2,3 FRED E. MYER2 AND WEN K. YANG2*

1University of Tennessee–Oak Ridge Graduate School of Biomedical Sciences, 2Biology Division, Oak Ridge National Laboratory, Oak Ridge, Tennessee 37831 and 3National Institute of Environmental Health Sciences, Research Triangle Park, North Carolina 27709, U.S.A.

(Accepted 28 November 1986)

SUMMARY

A 12.4 kbp HindIII chromosomal DNA fragment harbouring an apparently intact 9.2 kbp endogenous murine leukaemia virus (MuLV)-related proviral genome was isolated from an RFM/Un strain mouse by molecular cloning and designated pRFM # 6. Nucleotide sequence analysis revealed the following characteristic features in the pRFM # 6 provirus: a distinct 200 bp sequence in the long terminal repeat (LTR) mid-U3 region, a primer binding site for glutamine tRNA, a 3' pol region encoding an 'endonuclease' protein of 390 amino acids, and the mink cell focus-forming virus type-specific sequence at the 5' portion of the env gene. The 699 bp 5' LTR and 700 bp 3' LTR of pRFM # 6 provirus were identical except for three base changes in the U3 'enhancer' region. At the cell–provirus DNA junction, 4 bp direct repeats were present. The proviral genome was found at the same chromosomal DNA site in BALB/c, AKR, C3H, CBA and RFM strain mice, but not in NFS/N or C57BL/6 strain mice.

INTRODUCTION

All inbred strains of Mus musculus contain at least 20 to 50 copies (per haploid genome) of DNA sequences homologous to murine leukaemia viruses (MuLVs) (Chattopadhyay et al., 1974; also see review by Stoye & Coffin, 1985). These sequences are dispersed randomly throughout mouse chromosomal DNA and are transmitted vertically from parent to progeny. A small number (one to a few) of these sequences represent the loci for the biologically infectious xenotropic and ecotropic MuLVs and can be induced in vitro by a variety of chemicals, such as 5-iodo-2'-deoxyuridine and cycloheximide, to express infectious retroviruses (Lowy et al., 1971; Teich et al., 1973; Aaronson et al., 1974). These inducible loci have been studied extensively and the chromosomal locations of some have been mapped in certain strains of inbred mice (Kozak & Rowe, 1979, 1980). The inducibility and expression of the vast majority of the endogenous MuLV-related sequences, however, have not been well characterized. It is believed that some of these sequences are involved in the generation of recombinant dual-tropic mink cell cytopathic focus-forming (MCF) viruses which usually appear in spontaneous and ecotropic MuLV-induced lymphomas (Chattopadhyay et al., 1982; Khan et al., 1982; Khan, 1984). Most of these sequences apparently contain long terminal repeats (LTR) as well as gag, pol and env structural genes similar to those found in proviruses of all known infectious MuLVs (Dolberg et al., 1981; Rassart & Jolicoeur, 1982), although no extracellular retroviruses directly corresponding to the MuLV-related proviral sequences have been recognized. With the exception of one study (Bacheler, 1984), all molecular clones of mouse chromosomal non-ecotropic and non-xenotropic MuLV-related proviruses thus far reported have been incomplete or contain apparent deletions...
K. N. NIKBAKHT AND OTHERS

Fig. 1. Restriction endonuclease map of pRFM #6. DNA sequences of cellular origin are shown as thin lines flanking each LTR. The 5' flanking sequence was about 20 bp whereas the 3' flanking sequence was 3.2 kbp in size. The dark area in each LTR represents the novel 170 to 200 bp sequence segment present in the U3 region. Also shown is the 1.5 kbp \( BgII/BgII \) restriction fragment isolated from clone AL10 (Ou et al., 1983) and used as a \( gag \)-specific probe to distinguish 5' LTR-containing fragments from 3' LTR-containing fragments in the initial mapping analysis.

in the structural genes. Here we report the isolation of a mouse chromosomal DNA clone harbouring an apparently full-length MuLV-related proviral sequence that allows a detailed nucleotide sequence characterization of various provirus structural features in a single molecule.

METHODS

Preparation of chromosomal DNA and recombinant DNA clones. Chromosomal DNA was isolated from the spleen of a male RFM/Un mouse according to a standard procedure (Hoggan et al., 1982). Charon 9 lambda phage was used as the cloning vector (Benton & Davis, 1977; Boone et al., 1983; Ou et al., 1983) except that HindIII-digested DNA fragments from the RFM/Un mouse were used for insertion into the vector and an LTR probe from RFM/Un ecotropic MuLV (Liou et al., 1983; Nikbakht et al., 1985) was used for clone selection. A total of 20 clones were isolated which were subsequently transferred to pBR322. Further analysis showed that three of these chromosomal DNA clones contained two LTRs in a provirus-like structure. Clone pRFM #6 was selected for further characterization.

Restriction endonuclease mapping. All restriction enzymes were purchased from Bethesda Research Laboratories and New England Biolabs and digestions of cloned DNA were carried out according to the specifications recommended by the supplier. Restriction enzyme sites were determined by the partial digestion technique of Smith & Birnstiel (1976).

DNA gel blot analysis. Restriction DNA fragments were separated by horizontal electrophoresis using 0.7 or 1% submerged agarose gels and then transferred to nitrocellulose membranes according to the method described by Southern (1975). Methods for hybridization and autoradiography have been described (Yang et al., 1980; Ou et al., 1983). An LTR-specific probe was made by nick translation of an internal \( PstI/KpnI \) fragment isolated from the LTR of a proviral DNA clone of WN1802, the N-tropic MuLV derived from the BALB/c mouse (Boone et al., 1983). The ecotropic virus-specific probe was prepared by nick translation of a 310 bp \( AscI(SmaI) \) env fragment located 1.8 to 2.2 kbp from the 3' end of the ecotropic RFV proviral DNA clone (Chattopadhyay et al., 1980; Liou et al., 1983)

Nucleotide sequence analysis. Restriction fragments to be sequenced were subcloned into pUC9 (Ruther, 1982). DNA sequencing was performed according to the method of Maxam & Gilbert (1977).

RESULTS

Restriction enzyme map analysis

Fig. 1 shows the restriction endonuclease map of the 12.4 kbp HindIII DNA fragment cloned from the chromosomal DNA of the RFM/Un mouse and designated pRFM #6. The 9.2 kbp proviral genome of pRFM #6 contained several restriction sites commonly found in proviruses of endogenous ecotropic MuLV, xenotropic MuLV and MuLV-related sequences (Dolberg et al., 1981; Joseph, 1981; Chattopadhyay et al., 1982; Hoggan et al., 1982; Khan et al., 1982;
Fig. 2. Nucleotide sequences of the 5' and 3' LTRs and adjacent cellular and viral junctions in pRFM # 6. The proviral sequences shown are those of the plus strand, i.e. corresponding to retroviral genomic RNA. Also shown is the 5' LTR and primer binding site for minus strand DNA synthesis of RFV provirus (Liou et al., 1983). Using the nucleotide sequence of the 5' LTR of pRFM # 6 as the reference, dots (·) represent identical nucleotide bases and dashes (-) represent the absence of a nucleotide. The novel 170 to 200 bp sequence segment in the U3 region of pRFM # 6 LTRs is within brackets. Other sequences of structural or functional significance are boxed. Abbreviations: 'TATA' and 'CAT', RNA polymerase II consensus sequences; pA, polyadenylation signal; PBS, primer binding site for (−) strand synthesis; PU, putative primer site for (+) strand synthesis; IR, inverted repeat; DR, direct repeat.
Murine leukaemia virus-related genome

#6 nuc : AGGCTCTGCTGAGCACCACCACCACCTGTTCAACTAGCGCGCATCGCTGTGGATACCCGCACTCACGTAAAAGCGGCGG
      a.a. : ThrValLeuThrThrProThrAlaLeuValAspGlyIleAlaAlaAlaThrIleHisAlaAlaHisValGlyAlaAla
AKV nuc : AGGCTCTGCTGAGCACCACCACCACCTGTTCAACTAGCGCGCATCGCTGTGGATACCCGCACTCACGTAAAAGCGGCG
      a.a. : ThrValLeuThrThrProThrAlaLeuValAspGlyIleAlaAlaAlaThrIleHisAlaAlaHisValGlyAlaAla

#6 nuc : ACAACCCCTCCGGCCGGAACAGCATCAGGACCGACATAG
      a.a. : ThrThrProProAlaGlyThrAlaSerGlyProThr.
AKV nuc : ACAACCCCTCCGGCCGGAACAGCATCAGGACCGACATAG
      a.a. : ThrThrProProAlaGlyThrAlaSerGlyProThr.

AKV nuc : TAA
      a.a. : ***

Fig. 3. Comparison of nucleotide and predicted amino acid sequences of the 3' pol genes of pRFM # 6 and AKV. For sequencing, a 1.5 kbp SmaI/SmaI restriction fragment of pRFM # 6 (Fig. 1, map positions 5.1 to 6.7 kbp) including the 3' portion of pol and the 5' portion of env (Fig. 4) was subcloned into sequencing vector pUC9 (Ruther, 1982). Only nucleotide sequences of the putative endonuclease genes are shown here. The numbering system of Herr (1984) for the AKV sequence was adopted. The 5' Smal site of the cloned 1.5 kbp fragment corresponds to nucleotides 4579 to 4584 in AKV. The position of the NH2-terminal amino acid (i.e. Ile) in both polypeptides is according to Levin et al. (1984) and Herr (1984). This is encoded by an ATA codon which starts at nucleotide 4626 in AKV. Using the elucidated pRFM # 6 sequence as the reference, dots (.) and asterisks (*) respectively represent identical nucleotide bases and identical predicted amino acid residues found in AKV sequences; © © ©, = termination codon.

Roblin et al., 1982; Steffen et al., 1982; Liou et al., 1983; Ou et al., 1983; Herr, 1984; Nikbakht et al., 1985). These common sites (with their approximate distances in kbp from the 5' terminus of the 5' LTR of pRFM # 6) are: PstI(0.04/8.6)-SmaI(0.6/9.1)-KpnI(0.6/9.1) in the LTR, BglII(0.9) upstream of the gag gene, KpnI(3.4)-BamHI(3.8)-BamHI(4.2)-SalI(4.3)-XhoI(4.7)-BglII(4.7)-SmaI(5.2) in the pol gene and SmaI(6.8)-BglII(8.0) in the env gene. The EcoRI(6.9) site in the env gene is found in xenotropic MuLVs as well as in MuLV-related proviruses but not in ecotropic MuLV, whereas the BamHI(6.3) site in the env gene is found only in endogenous MuLV-related and MCF proviral clones.

Nucleotide sequences of LTRs, primer binding sites and cell–provirus junction

The 5' and 3' LTR structures of pRFM # 6 as well as the LTR of RFM/Un mouse endogenous ecotropic MuLV, RFV (Liou et al., 1983) are shown in Fig. 2. The 5' LTR was 699 bp long and the 3' LTR 700 bp long; their nucleotide sequences were identical except for one base deletion/addition and two base transitions, all within the putative U3 'enhancer' area. The two transcriptional signals, CCAAC or the 'CAT' box and the TATAAAAA or the 'TATA', box are identified in the U3 region. The 'CAT' box of pRFM # 6 LTR (i.e. CCAAC compared to CCAAT) was six nucleotides farther upstream from the cap site than that of RFV because of a CAGCCC duplication on the 5' side of the TATA box, a feature also found in LTRs of some other MuLV-related proviral clones (Khan & Martin, 1983; Ou et al., 1983). Greater than 90% sequence homology was observed in the U5 and R regions of the LTRs of pRFM # 6 and RFV proviruses, whereas the U3 region was distinctly different in the two proviral genomes. The presence of a unique 200 bp internal sequence (bracketed in Fig. 2 in the mid-U3 region of pRFM # 6 had made this LTR correspondingly larger than that of RFV because of a CAGCCC duplication on the 5' side of the TATA box, a feature also found in LTRs of some other MuLV-related proviral clones (Khan & Martin, 1983; Ou et al., 1983). Greater than 90% sequence homology was observed in the U5 and R regions of the LTRs of pRFM # 6 and RFV proviruses, whereas the U3 region was distinctly different in the two proviral genomes. The presence of a unique 200 bp internal sequence (bracketed in Fig. 2 in the mid-U3 region of pRFM # 6 had made this LTR correspondingly larger than that of RFV provirus. As a result, the proviral genome of approximately 9.2 kbp in pRFM # 6 is larger than the approximately 8.8 kbp proviral genomes of known infectious type C viruses. The remainder of the U3 region in pRFM # 6 shows only 77% homology to the corresponding RFV LTR.

The stretch of 18 nucleotides immediately 3' to the U5 of the 5' LTR in pRFM # 6 (Fig. 2), the presumed tRNA primer binding site for the initiation of minus strand DNA synthesis, TGGAGTCCACCCAGAT, is a perfect complementary sequence to the 3' terminal 18 nucleotides of the glutamine isoacceptor tRNA (Yang et al., 1983). This is distinctly different from the proline tRNA binding site found in most infectious mammalian type C retroviruses (Waters, 1975; Harada et al., 1979). In contrast, the putative site for the plus strand DNA
## Fig. 4. Comparison of nucleotide and predicted amino acid sequences of the corresponding 5' portion of env genes in pRFM # 6, AKR MCF247 MuLV (Khan, 1984) and NFS xenotropic MuLV (Khan, 1984).

This portion of the env gene is located at the 3' end (between BamHI site at 6.25 and SmaI site at 6.7 kbp in Fig. 1) of the 1.5 kbp SmaI/SmaI fragment described in the legend of Fig. 3. Using the elucidated pRFM #6 sequence as the reference, dots (.) and asterisks (*) respectively represent identical nucleotide bases and identical predicted amino acid residues found in the AKR MCF247 and the NFS xenotropic MuLV sequences. The underlined sequence in pRFM # 6 indicates the region where marked similarity to, or differences from, the two MuLV types is noted.

synthesis (Gilboa et al., 1979) of pRFM #6 was AGAAAGAGGGGGG, which shows considerable homology to those found in AKV and RFV (Liou et al., 1983; Herr, 1984) and other type C mammalian retroviruses (Chen & Barker, 1984).

The pRFM #6 provirus was bounded by inverted repeats of 11 nucleotides, i.e. TGAAAGACCC at the 5' end and GGGGTCTTTC at the 3' end (Fig. 2) and thus is similar to the integrated proviruses of MuLVs, which show the loss of AA/TT from the canonical AATGAAAGACCCC/GGGGTCTTTCATT inverted repeats of unintegrated linear viral DNA intermediates (Chen & Barker, 1984). Furthermore, the inverted repeats at the termini of pRFM #6 provirus were linked to direct repeats of four nucleotides, CCAA/CCAA, similar to the 4 bp sequence duplication found at cellular integration sites of exogenous retroviruses.
The 3′ sequences of the pol gene

Studies with mutant MuLVs have implied that the 3′ portion of the pol gene encodes a polypeptide (mol. wt. 40K to 45K) with the presumed endonuclease function required for integration of proviral DNA (Donehower & Varmus, 1984; Schwartzberg et al., 1984; Levin et al., 1984). Therefore, we have sequenced the 3′ region of the pol gene in pRFM # 6 and compared the predicted amino acid sequence with the corresponding part of AKV (Herr, 1984). As shown in Fig. 3, the open reading frame of the putative endonuclease gene in pRFM # 6 is 1170 nucleotides long and is terminated by an amber (TAG) codon, whereas that of AKV is 1215 nucleotides long and terminated by an ochre (TAA) codon. As a result, the deduced polypeptide sequence for pRFM # 6 endonuclease is 15 amino acids shorter at the carboxy terminus than that of AKV. The remaining portions of the AKV and pRFM # 6 endonucleases were 86% homologous in amino acid sequence and 80% homologous in nucleotide sequence.

Characterization of env gene sequences

Since MuLVs of different host range types (i.e. ecotropic, xenotropic and dual-tropic) may be distinguished by sequences present in specific regions of the env gene, we examined this property of the pRFM # 6 proviral genome by hybridization as well as by nucleotide sequence analysis. The pRFM # 6 provirus showed no hybridization to an ecotropic type-specific env probe (data not shown). It has been demonstrated that the MCF-type and the xenotropic-type MuLVs differ significantly in nucleotide sequence at the 5′ portion of env gene (Khan, 1984; O’Neill et al., 1985). Nucleotide sequences elucidated for this env gene portion of pRFM # 6 (Fig. 4) are clearly more related to those of the AKR MCF247 (Khan, 1984) than to NFS/N xenotropic MuLV (Khan, 1984) or NZB9-1 xenotropic MuLV (O’Neill et al., 1985). In the particular 100 nucleotide segment bounded by BstNI and XmaI sites, where the MCF-type and the xenotropic-type differ the most, pRFM # 6 is identical with the AKR MCF247 except for a G to A base transition (and hence the absence of the BstNI site), whereas it differs from the NFS/N xenotropic MuLV by 21 base changes and a 12 nucleotide sequence deletion/addition.

Cellular DNA site of pRFM # 6 proviral integration in other mouse strains

Gel blot analysis revealed the presence of cellular repetitive sequences in the immediate 3′ flanking region of pRFM # 6 provirus. However, unique cellular sequences were found at the 3′ side of the Smal site (map position 10.6, Fig. 1) and therefore were used as probes to analyse the pRFM # 6 proviral integration site in RFM/Un and six other laboratory mouse strains. As shown in Fig. 5(a), the molecular probe prepared from the 1·2 kbp EcoRI/HindIII fragment (map positions 11.2 to 12.4) detected a single 12·4 kbp HindIII fragment in chromosomal DNA preparations from RFM/Un, BALB/c, AKR, C3H and CBA mice and a single 3·2 kbp HindIII fragment in those from NFS/N and C57BL/6 mice. The 12·4 kbp is precisely the size of the pRFM # 6 DNA clone whereas 3·2 kbp is the expected size of 12·4 kbp minus the proviral length of 9·2 kbp. In addition, the molecular probe prepared from the 0·6 kbp SamI/EcoRI fragment of pRFM # 6 clone (map positions 10·6 to 11·2), despite having a small amount of repetitive sequences, detected the predicted 4·4 kbp DNA band in EcoRI-digested chromosomal DNAs of RFM/Un, BALB/c and AKR mice but not in those of C57BL/6 or NFS/N mice (Fig. 5b). These results indicate that BALB/c, AKR, C3H and CBA mice also possess the pRFM # 6 locus (i.e. specific cellular DNA site plus the integrated MuLV-related provirus) whereas no provirus nor an LTR is present in the same DNA site of NFS/N or C57BL/6 mice.

DISCUSSION

Despite the ubiquity of endogenous MuLV-related proviral sequences in chromosomal DNA of laboratory mice, isolation of these sequences in intact forms has been hindered in several cases either due to the choice of restriction enzymes for molecular cloning or by deletions of various sizes in the structural genes (Khan et al., 1982; Roblin et al., 1982; Steffen et al., 1982; Ou et al., 1983; Nikbakht et al., 1985). The present study of pRFM # 6 was possible because some MuLV-related sequences in chromosomal DNA possess no internal HindIII restriction enzyme sites (Ou et al., 1983).
Fig. 5. Southern gel blot analysis of the pRFM # 6 provirus locus in various mouse strains. Restriction enzyme-digested mouse liver DNA preparations (5 to 10 μg per lane) were separated in 0.7% agarose gel by horizontal electrophoresis. (a) HindIII-restricted mouse DNA fragments revealed by a sequence probe prepared from a 1.2 kbp EcoRI/HindIII fragment of pRFM # 6 (map positions 11.2 to 12.4, Fig. 1). (b) EcoRI-restricted mouse DNA fragments detected by a sequence probe from a 0.6 kbp SmaI/EcoRI fragment of pRFM # 6 (map positions 10.7 to 11.3, Fig. 1). This 0.6 kbp cellular flanking sequence contained a small portion of a mouse genomic repetitive sequence and hence gave a high background in the gel blot. The arrow indicates a distinct 7.0 kbp hybridizing band in the C57BL/6 and NFS/N mouse DNA samples. pRFM # 6 clone DNA digested with HindIII or EcoRI was included as a control. The molecular probes, prepared by nick translation, had a specific activity of approx. 3 × 10^8 c.p.m./μg DNA.
The results of our nucleotide sequence analysis of LTRs, tRNA binding site, polypurine primer site and 4 bp duplicated junctions clearly imply that the endogenous MuLV-related proviral genome in pRFM # 6 is similar to horizontally transmitted and exogenously acquired proviral genomes with respect to replicative and integration mechanisms (for review, see Varmus & Swanstrom, 1985). Also, the apparently intact sequences of these characteristic features, including the virtually identical 3' and 5' LTR sequences, would suggest that the pRFM # 6 MuLV-related proviral genome either was acquired rather recently by the mouse germ line or has been subjected to minimal changes in the mouse genome. Furthermore, the potential transcriptional signals in the pRFM # 6 LTRs are consistent with recent findings of mRNA expression from endogenous MuLV-related proviruses in the mouse (Levy et al., 1985; our unpublished observation).

On the other hand, the pRFM # 6 proviral genome also has some structural properties that distinguish the MuLV-related sequences from the known endogenous ecotropic and xenotropic MuLVs. First, with the apparent insertion of a distinct 200 bp sequence segment between the 'CAT' box and the 'enhancer' sequence in the mid-U3 region (Khan & Martin, 1983; Ou et al., 1983), the LTR of pRFM # 6 provirus is larger than those of infectious MuLV proviruses. Second, sequencing studies of the tRNA primer binding sites in many MuLV-related clones imply the use of glutamine tRNA as the primer for minus strand DNA synthesis, in contrast to proline tRNA in infectious ecotropic MuLV (Ou et al., 1983; Nikbakht et al., 1985). This is also true for the pRFM # 6 provirus. Third, the 5' nucleotide sequence of env in pRFM # 6 is evidently of the MCF type but not of the xenotropic MuLV type. Thus, the characteristic env sequences that specify the MCF recombinant MuLV type may be derived from distinct MuLV-related proviral structures in the mouse chromosome. However, it remains to be determined whether the MCF type-specificity is associated with all or only some of the MuLV-related proviruses that possess the distinct LTR and primer tRNA binding features.

One interesting question is, therefore, when this pRFM # 6 provirus carrying the MCF specificity was incorporated into the germ line of RFl/Un strain mice. Our results indicate that the pRFM # 6 locus is also present in BALB/c, AKR, C3H and CBA strain mice. The BALB/c, C3H and CBA strains are derived from the same origin, i.e. Bagg albino. The RFl/Un strain may be related to the BALB/c and the C3H strains, since they carry an apparently common ecotropic MuLV locus, emv-1, on chromosome 5 (our unpublished results). Although AKR mice could be of distinct lineage, this and the RFl/Un strain were initially established in the same laboratory. On the other hand, the distantly related NFS/N and C57BL/6 strain mice harbour no provirus at this chromosomal DNA site. With a previous observation (Ou et al., 1983), this suggests that incorporation of MuLV-related proviruses into the mouse germ line may have occurred after speciation of Mus musculus domesticus.

We thank Dr Raymond R. O'Neill of the National Institute of Allergy and Infectious Diseases, National Institutes of Health, for discussion with us on type-specific env sequences. We also thank Ms Neva Hair for typing the manuscript. This research has been jointly supported by the National Institute of Environmental Health Sciences (Public Health Service grant ES-4-0118), the National Cancer Institute (Public Health Service grants CA30308 and CA09104), and the Office of Health and Environmental Research, U.S. Department of Energy, under contract DE-AC05-840R21400 with Martin Marietta Energy Systems, Inc.

REFERENCES


(Received 28 July 1986)