JOURNAL OF VIROLOGY, July 1997, p. 5652–5657 Vol. 71, No. 7
0022-538X/97/$04.00+0

Cloning of a New Murine Endogenous Retrovirus, MuERV-L, with Strong Similarity to the Human HERV-L Element and with a *gag* Coding Sequence Closely Related to the *Fv1* Restriction Gene

LAURENCE BÉNIT,[1] NATHALIE DE PARSEVAL,[1] JEAN-FRANÇOIS CASELLA,[1] ISABELLE CALLEBAUT,[2] AGNÈS CORDONNIER,[1]† AND THIERRY HEIDMANN[1]*

Unité de Physicochimie et Pharmacologie des Macromolécules Biologiques, CNRS URA 147, Institut Gustave Roussy, 94805 Villejuif Cedex,[1] and Systèmes Moléculaires et Biologie Structurale, LMCP, CNRS URA 09, UP6-UP7, 75252 Paris Cedex 05,[2] France

Received 12 December 1996/Accepted 20 March 1997

We had previously identified a new family of human endogenous retrovirus-like elements, the HERV-L elements (human endogenous retrovirus with leucine tRNA primer), whose *pol* gene was most closely related to that of the foamy viruses. HERV-L *pol*-related sequences were also detected in other mammalian species. The recent cloning of the mouse *Fv1* gene (S. Best, P. Le Tissier, G. Towers, and J. P. Stoye, Nature 382:826–829, 1996) has shed light on another HERV-L domain—identified as a *gag* gene based on its location within the provirus—which was found to be 60% identical, at the nucleotide level, to the *Fv1* open reading frame (ORF). We have now cloned the murine homolog of HERV-L which, in contrast to HERV-L, displays fully open reading frames in the *gag* and *pol* genes. Its predicted Gag protein shares 43% identity with the Fv1 ORF product. Moreover, the characteristic major homology region of the capsid subdomain can be identified within both proteins, thus strongly emphasizing the *gag*-like origin of *Fv1*.

We had previously identified a new family of human endogenous retrovirus-like elements, the HERV-L elements (for human endogenous retrovirus with leucine tRNA primer), which displayed a foamy virus-related *pol* sequence and had expanded in the genome of primates to a level of up to 200 copies (4). A zoo blot hybridization under rather stringent conditions with a *pol*-specific probe had revealed the presence of a limited number of related sequences in most mammals, with a burst in mice—not observed in the rat—of up to 100 to 200 copies. The cloned HERV-L element disclosed many stop codons, but fragmented open reading frames (ORFs) could be identified for the *pol* products, including reverse transcriptase, RNase H, and integrase, and for a dUTPase. The *gag* region of HERV-L was recently found as the sequence the most closely related to the newly cloned murine *Fv1* gene (1, 2). The *Fv1* (Friend virus susceptibility 1) gene controls the replication of murine leukemia retroviruses and prevents disease in mice infected with these viruses (9, 11, 17). This sequence similarity was highly suggestive of a Gag-like structure for the Fv1 product and indicated that the *Fv1* gene could have a retroviral origin. However, due to many stop codons in the cloned HERV-L element, the identification of the *gag* region was questioned, since it was mostly based on its position in the genome. In addition, the *Fv1* gene is present only in the mouse (1), and a likely progenitor for this gene would be rather expected to be a murine sequence. These questions can now be addressed.

Cloning of MuERV-L and comparison with HERV-L. Since our 1995 publication (4), we have isolated and characterized the murine homolog of HERV-L, which we have named MuERV-L. This new murine endogenous retrovirus was cloned from a BALB/c mouse genomic library, by using as a probe a 360-bp fragment from the HERV-L *pol* gene and following standard procedures as described in reference 4. A full-length proviral element, as well as phage clones containing only part of other copies of MuERV-L, was entirely sequenced. As illustrated in Fig. 1, a dot matrix comparison of HERV-L and MuERV-L using the COMPARE program of the Genetics Computer Group package shows that both sequences are closely related. The long terminal repeats (LTRs) are the most diverged sequences between MuERV-L and HERV-L, with only 40% similarity, while the coding regions are 74% identical. At the amino acid level, a striking feature of the randomly cloned MuERV-L element is that it harbors an almost full coding potential. A first ORF (from nucleotides [nt] 538 to 2283) most probably corresponds to the *gag* domain of the element (see next sections), whereas the second ORF was interrupted by a frameshift located between nt 4100 and 4250. However, sequencing of the corresponding domain of other partial MuERV-L elements revealed an 8-bp deletion in the cloned full-length copy, the reintroduction of which restores a complete ORF for *pol* extending from nt 2284 to 5831 (see ORF map in Fig. 1 and position of the 8-bp insertion). MuERV-L and HERV-L display a similar overall organization, including the characteristic presence of a dUTPase gene 3′ to *pol* and the absence of an *env* gene.

Sequence analysis of MuERV-L. The MuERV-L proviral sequence is 6,471 nt long (Fig. 2). The 5′ and 3′ LTRs of MuERV-L are 98% identical over 493 bp. They present the usual features of retroviral LTRs, i.e., they are bordered by

* Corresponding author. Mailing address: Unité de Physicochimie et Pharmacologie des Macromolécules Biologiques, CNRS URA 147, Institut Gustave Roussy, 39 rue Camille Desmoulins, 94805 Villejuif Cedex, France. Phone: 33-1.42.11.49.70. Fax: 33-1.42.11.52.76. E-mail: heidmann@igr.fr.

† Present address: Cancérogenèse et Mutagenèse Moléculaire et Structurale, CNRS UPR 9003, Ecole Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France.

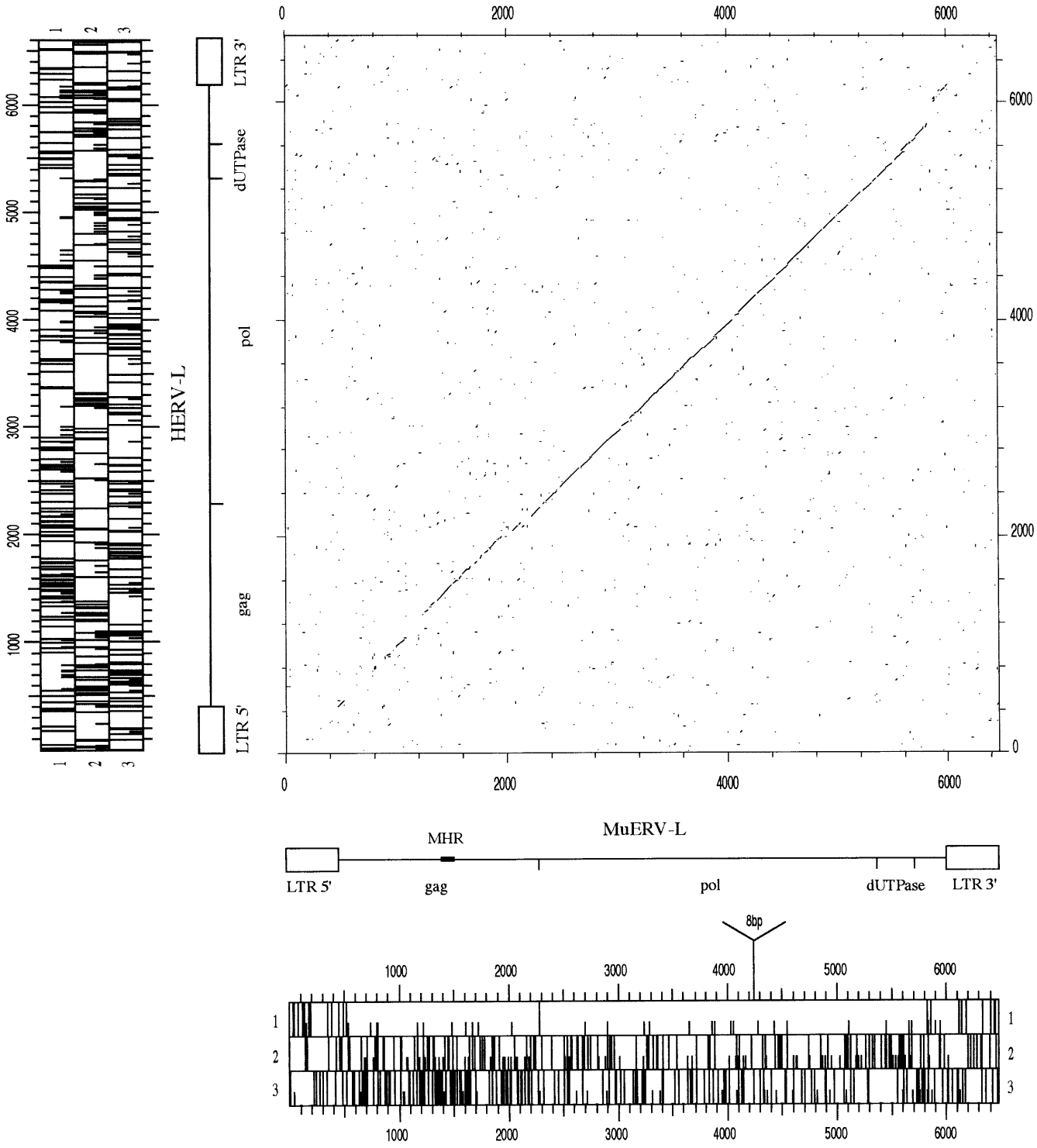


FIG. 1. Structure of the MuERV-L provirus and comparison with HERV-L: dot matrix comparison between MuERV-L and HERV-L and ORF map of the HERV-L provirus (left) and the MuERV-L provirus (bottom). MHR, major homology region.

short inverted repeats (TGTA. . .TACA) and contain a presumptive TATA box and a polyadenylation signal. The identification of a CAAT box is more uncertain. Two degenerated sequences, 30 and 65 bp upstream of the TATA box, may represent imperfect CAAT boxes. As for HERV-L, a putative tRNA primer-binding site complementary to the 3′ end of a mouse leucine tRNA (16) can be identified downstream of the 5′ LTR (5 nt from the inverted terminal repeat of the LTR, as observed for another human endogenous retrovirus [15]), as well as a polypurine track close to the 3′ LTR.

Unlike HERV-L, which contains many stop codons, the predicted MuERV-L translation shows a long ORF (as can be

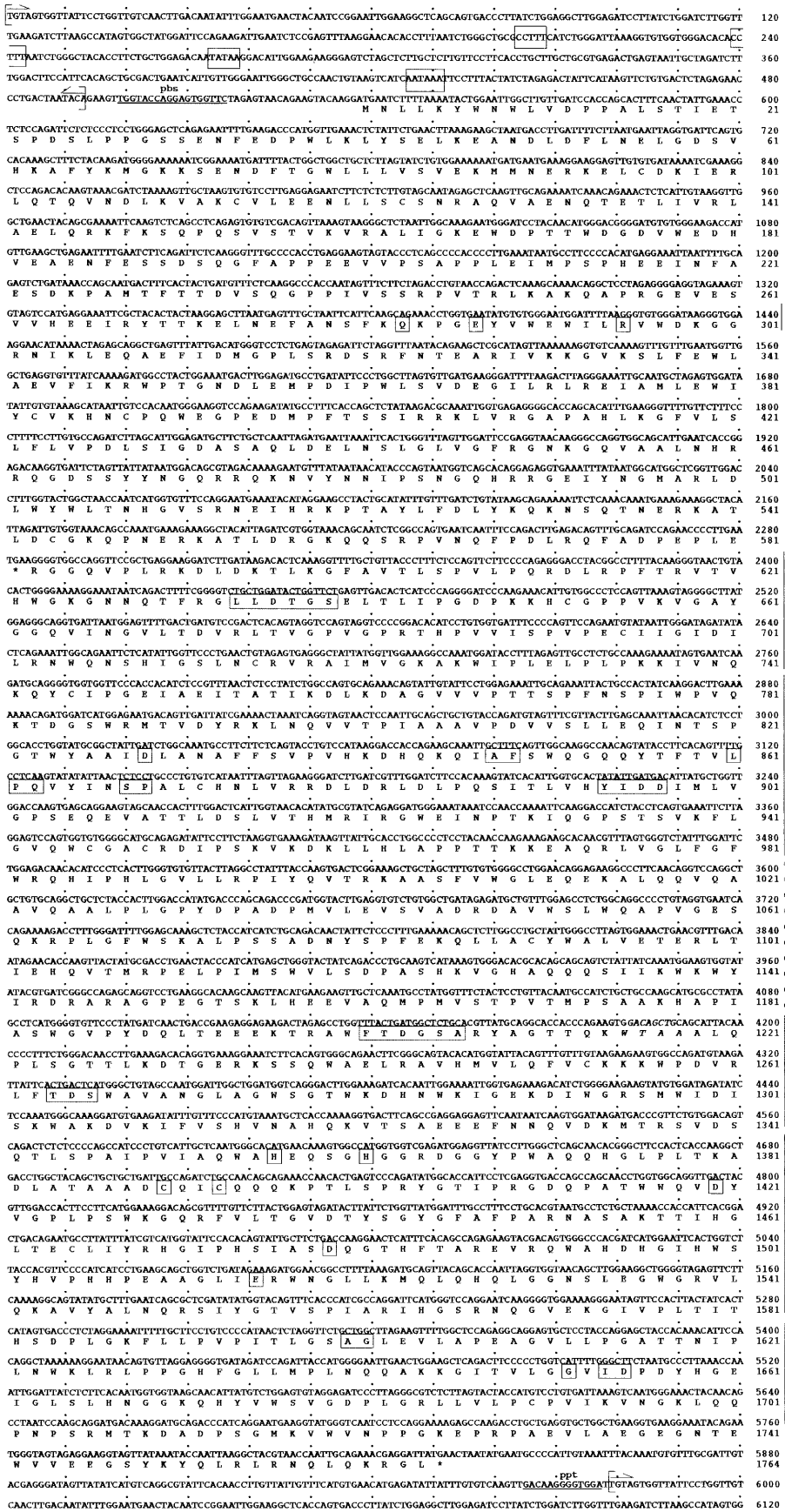
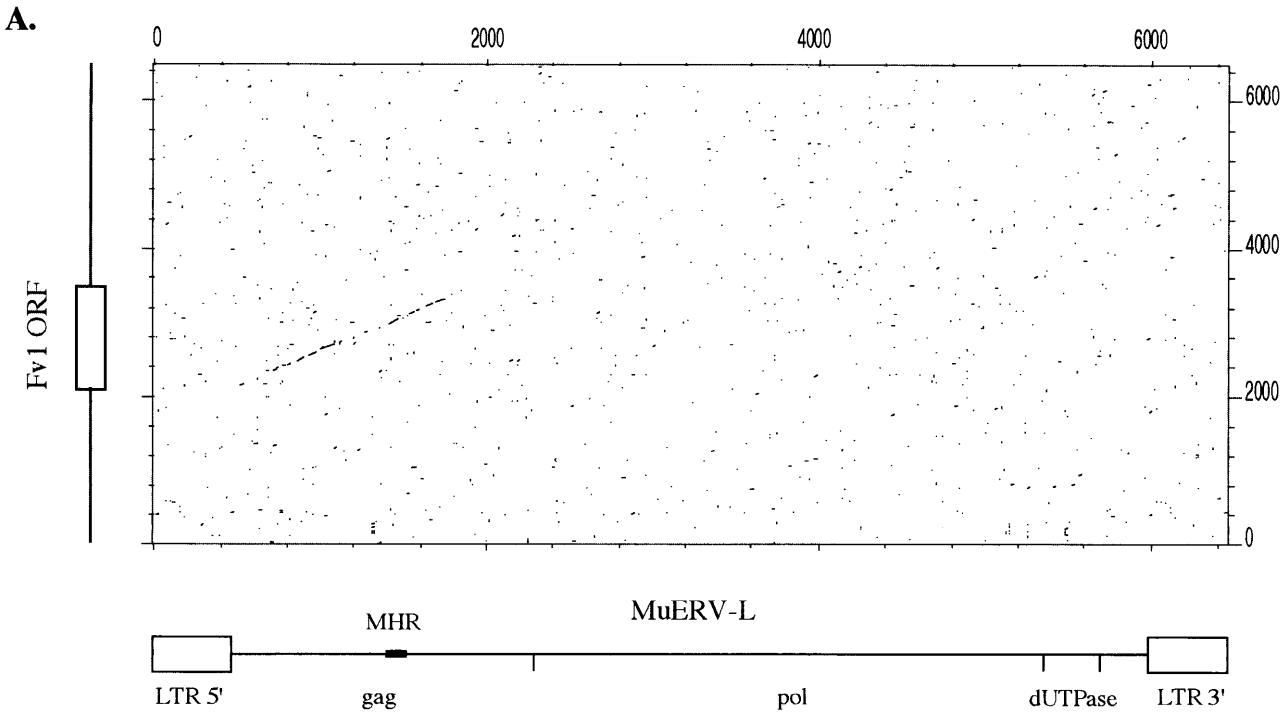


FIG. 2. Nucleotide sequence of MuERV-L proviral DNA and predicted translation products (single-letter amino acid code; asterisk, stop codon). LTRs are enclosed by brackets, and the small inverted termini are overlined with arrows. The two degenerated CAAT boxes are boxed, as are the TATA box and the polyadenylation signal. The primer-binding site (pbs) and polypurine tract (ppt) are underlined. The 8 nt which are deleted in the cloned full-length element are indicated in italics (positions 4182 to 4189). The different predicted proteins are indicated. The conserved residues (see text) within the protease, reverse transcriptase, RNase H, integrase, and dUTPase are boxed, as are the three conserved residues within the MHR motif.

seen on the ORF map in Fig. 1) except for a stop codon at position 2281 which most probably corresponds to the end of Gag. In this respect, regions downstream of the *gag* gene would be translated by a readthrough suppression mechanism as observed for some retroviruses, such as the mammalian type C retroviruses (reviewed in reference 3). Coding capacities of the *gag* gene were confirmed by the size of a recombinant protein that we made in bacteria (75 kDa [data not shown]). With the



B.

```
Fv1         1 MNFPRALAGFSSWLFK.PELAEDSPDNDSPDNDTVNPWRELLQKINVADLPDSSFSSGKELNDSVYHTFEHFCKIRDYDAVGELLLAFLDKVTKERDQFR
MuERV-L     1. ...MNLLKYWNWLVDPPALSTIETSPDSLPPGSSENFEDPWLKLYSELKEANDLDFLNELGDSVHKAFYKMGKKSENDFTGWLLLVSVEKMMNERKELC
Fv1       100 DEISQLRMHINDLKASKCVLGETLLSYRHRIEVGEKQTEALIVRLADVQSQVMCQPARKVSADKVRALIGKEWDPVTWDGDVWED......IDSEGSEE.
MuERV-L    97 DKIERLQTQVNDLKVAKCVLEENLLSCSNRAQVAENQTETLIVRLAELQRKFKSQP.QSVSTVKVRALIGKEWDPTTWDGDVWEDHVEAENFESSDSQGF
Fv1       193 AELPTVLAS.......PSLSEESGYALSKERTQQDKADAPQIQSSTSLVTSEPVTRPKS.......LSDLTSQKHRHTNHELNSLAHSNRQKAKEHARKW
MuERV-L   196 APPEEVVPSAPPLEIMPSPHEEINFAESDKPAMTFTTDVSQ...GPPIVSSRPVTRLKAKQAPRGEVESVVHEEIRYTTKELNEFANSFKQKPGEYVWEW
Fv1       279 ILRVWDNGGRLTILDQIEFLSLGPLSLDSEFNVIARTVEDNGVKSLFDWLAEAWVQRWPTTRELQSPDTLEWYSIEDGIERLRELGMIEWL.CVKATCPQ
MuERV-L   293 ILRVWDKGGRNIKLEQAEFIDMGPLSRDSRFNTEARIV.KKGVKSLFEWLAEVFIKRWPTGNDLEMPD.IPWLSVDEGILRLREIAMLEWIYCVKHNCPQ
Fv1       378 WRGPEDVPITRAMRITFVRETRETWKSFVFSLLCIKDITVGSVAAQLHDLIELSLKPTAAGLTSVGSVGVLSLSPWKHQSNS*
MuERV-L   391 WEGPEDMPFTSSIRRKLVRGAPAHLKGFVLSLFLVPDLSIGDASAQLDELNSLGL....VGFRG.NKGQVAALNHRRQGDSSNKGQVAALNHRRQGDSSY
MuERV-L   469 YNGQRRQKNVYNNIPSNGQHRRGEIYNGMARLDLWYWLTNHGVSRNEIHRKPTAYLFDLYKQKNSQTNERKATLDCGKQPNERKATLDRGKQQSRPVNQF
MuERV-L   569 PDLRQFADPEPLE*
```

FIG. 3. (A) Dot matrix comparison between MuERV-L and the *Fv1* resistance gene. (B) Amino acid alignment between the MuERV-L *gag* coding region and the Fv1 ORF. The MHR is underlined in each sequence; the vertical bars indicate identity.

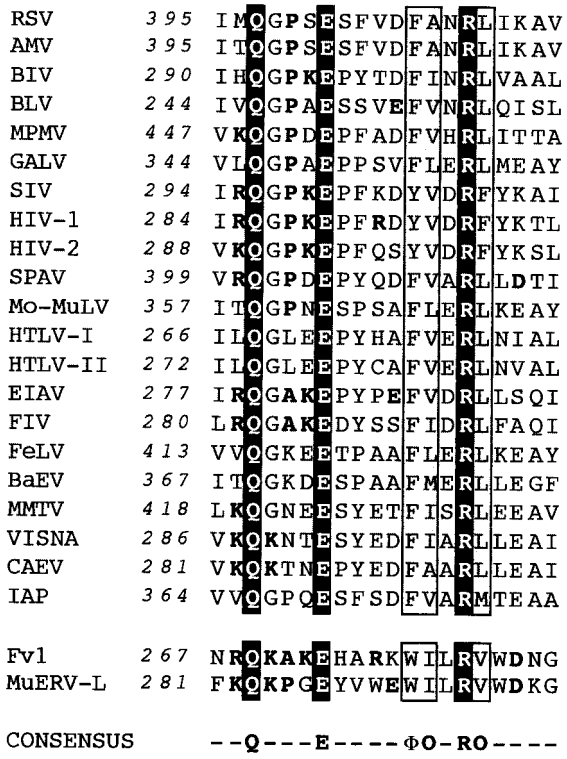


FIG. 4. Amino acid alignment of the MHR domain of a series of retroviruses (5) and a murine retrotransposon (IAP-MIA14) (13) of MuERV-L and Fv1 (1). The numbers indicate the position of the first residue of the MHR. The invariant residues Q, E, and R are indicated within black boxes, and the conserved hydrophobic residues are indicated within shaded boxes; Fv1 or MuERV-L amino acids also found in other MHR are printed in boldface. The consensus sequence is indicated, with Φ for aromatic and O for hydrophobic amino acids. RSV, Rous sarcoma virus; AMV, avian myeloblastosis virus; BIV, bovine immunodeficiency-like virus; BLV, bovine leukemia virus; MPMV, Mason-Pfizer monkey virus; GALV, gibbon ape leukemia virus; SIV, simian immunodeficiency virus; HIV-1 and -2, human immunodeficiency virus types 1 and 2; SPAV, sheep pulmonary adenomatosis virus; Mo-MuLV, Moloney murine leukemia virus; HTLV-I and -II, human T-cell leukemia viruses types 1 and 2; EIAV, equine infectious anemia virus; FIV, feline immunodeficiency virus; FeLV, feline leukemia virus; BaEV, baboon endogenous retrovirus; MMTV, mouse mammary tumor virus; CAEV, caprine arthritis encephalitis virus.

help of the FASTA computer program, the *pol* origin of the region from residues 583 to 1764 was ascertained by the relative positions of specific amino acids shared by all retroviral enzymatic proteins. These conserved residues and/or motifs are as follows (Fig. 2): for protease, the L-L-D-T-G-S motif (18); for reverse transcriptase, D and A-F within the third box of homology of reverse transcriptases (8, 21), L-P-Q and S-P within the fourth box, and Y-I-D-D within the fifth box; for RNase H, the F-T-D-G-S-A motif as well as a T-D-S motif (8); and in integrase, the imperfect zinc finger H-X_4-H-X_{22-32}-C-X_2-C (the histidine residues are separated in MuERV-L by four residues instead of three) and the specific D-X_{55-60}-D-X_{35}-E motif (8, 10). Therefore, MuERV-L potentially encodes all the usual retroviral enzymes, with the protease in the same frame as the *pol* products. As for HERV-L, MuERV-L encodes, at the carboxy-terminal end of the *pol* gene, a dUTPase-like protein with the A-G and G-X-I-D conserved residues (4, 12).

Homology between the Gag MuERV-L protein and the *Fv1* ORF product. Expectedly, as for HERV-L, the *gag* MuERV-L region displays similarities with the *Fv1* locus (Fig. 3). The domain of similarity coincides with the first three-fourths of MuERV-L Gag and covers the whole *Fv1* coding sequence product, where 43% amino acids are identical (Fig. 3B). The carboxy-terminal region of the *gag* product of MuERV-L that could include a nucleocapsid subdomain is not found in the Fv1 ORF product. A matrix subdomain could not be ascertained, as retroelements have only limited homology in this domain at the primary sequence level. However, in the capsid (CA) subdomain, the MuERV-L Gag protein and Fv1 ORF product display the characteristic major homology region (MHR) domain located at the carboxy terminus of the CA subdomain of most retroviruses (5, 14, 19) with the three absolutely conserved residues of this motif (Fig. 4), as well as the exact spacing between them (Q-X_3-E-X_7-R). These hydrophilic residues are embedded within hydrophobic residues, as observed for other Gag proteins (Fig. 4). Finally, further analysis using a combination of one-dimensional search methods and hydrophobic cluster analysis (6, 20) and comparison with the predicted structure of the human immunodeficiency virus type 1 CA obtained by crystallographic data (14) indicates structural similarities 5′ to the MHR, with a series of predicted α-helices that should form a coiled-coil structure (1a). These data strongly emphasize the *gag*-like origin of *Fv1* (1, 7). No other homology could be identified out of the *Fv1* coding region, including between the predicted promoter sequence of *Fv1* and the MuERV-L LTRs.

HERV-L-related sequences are found at a low copy number in all mammalian species, except in primates and in the mouse where they have been amplified up to 100 to 200 copies (4). In contrast to HERV-L, amplification of MuERV-L sequences should be recent, as suggested by the uninterrupted ORFs of the *gag* and *pol* genes in the cloned elements and the almost fully identical LTRs (98%), as well as the conservation among all the genomic copies of a series of restriction sites that can be tested by Southern blot analysis (reference 4 and data not shown). Moreover, a similar burst is absent in the rat genome. Therefore, MuERV-L should correspond to a recently amplified functional mouse endogenous element. As proposed by Best et al. (1), the homology between HERV-L and *Fv1*, the unusual structure of *Fv1* mRNA (a unique large exon), and its absence in a very closely related species, i.e., in the rat genome, lead to the hypothesis that the *Fv1* gene is likely to have evolved from an endogenous retrovirus. The newly identified MuERV-L endogenous retrovirus provides a murine retrovirus-like element with an appropriate *gag* coding sequence for the generation of this *Fv1* resistance gene. Distribution of the MuERV-L element among wild mice and functionality of the cloned copy are under investigation.

Nucleotide sequence accession number. The MuERV-L sequence has been entered in the EMBL database under the no. Y12713.

We are grateful to M. Lazar for providing the mouse BALB/c genomic library and C. Lavialle for critical reading of the manuscript.

This work was supported by the CNRS (ACC-SV3) and by grants from the ARC (contrat 6552 to T.H.) and Rhone-Poulenc Rorer (Bioavenir fellowships to L.B., N.P., and J.-F.C.).

REFERENCES

1. **Best, S., P. Le Tissier, G. Towers, and J. P. Stoye.** 1996. Positional cloning of the mouse retrovirus restriction gene *Fv1*. Nature **382:**826–829.
1a. **Callebaut, I.** Unpublished data.
2. **Coffin, J. M.** 1996. Retrovirus restriction revealed. Nature **382:**762–763.
3. **Coffin, J. M.** 1996. Retroviridae: the viruses and their replication, p. 1767–1847. *In* B. N. Fields, D. M. Knipe, P. M. Howley et al. (ed.), Fields virology 3rd ed. Lippincott-Raven Publishers, Philadelphia, Pa.
4. **Cordonnier, A., J.-F. Casella, and T. Heidmann.** 1995. Isolation of novel human endogenous retrovirus-like elements with foamy virus-related *pol* sequence. J. Virol. **69:**5890–5897.

5. **Craven, R. C., A. E. Leure-duPree, R. A. J. Weldon, and J. W. Wills.** 1995. Genetic analysis of the major homology region of the Rous sarcoma virus Gag protein. J. Virol. **69:**4213–4227.
6. **Gaboriaud, C., V. Bissery, T. Benchetrit, and J. P. Mornon.** 1987. Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. FEBS Lett. **224:**149–155.
7. **Goff, S. P.** 1996. Operating under a Gag order: a block against incoming virus by the *Fv1* gene. Cell **86:**691–693.
8. **Johnson, M. S., M. A. McClure, D.-F. Feng, J. Gray, and R. F. Doolittle.** 1986. Computer analysis of retroviral *pol* genes: assignment of enzymatic functions to specific sequences and homologies with non viral enzymes. Proc. Natl. Acad. Sci. USA **83:**7648–7652.
9. **Jolicoeur, P.** 1979. The *Fv1* gene of the mouse and its control of murine leukemia virus replication. Curr. Top. Microbiol. Immunol. **86:**67–122.
10. **Kulkoski, J., K. S. Jones, R. A. Katz, J. P. G. Mack, and A. M. Skalka.** 1992. Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. Mol. Cell. Biol. **12:**2331–2338.
11. **Lilly, F., and T. Pincus.** 1973. Genetic control of murine viral leukemogenesis. Adv. Cancer Res. **17:**231–277.
12. **McGeoch, D. C.** 1990. Protein sequence comparisons show that the "pseudoproteases" encoded by poxviruses and certain retroviruses belong to the deoxyuridine triphosphatase family. Nucleic Acids Res. **18:**4105–4110.
13. **Mietz, J. A., Z. Grossman, K. K. Lueders, and E. L. Kuff.** 1987. Nucleotide sequence of a complete mouse intracisternal A-particle genome: relationship to known aspects of particle assembly and function. J. Virol. **61:**3020–3029.
14. **Momany, C., L. C. Kovari, A. J. Prongay, W. Keller, R. K. Gitti, B. M. Lee, A. E. Gorbalenya, L. Tong, J. McClure, L. S. Ehrlich, M. F. Summers, C. Carter, and M. G. Rossmann.** 1996. Crystal structure of dimeric HIV-1 capsid protein. Nat. Struct. Biol. **3:**763–770.
15. **Repaske, R., P. Steele, R. O'Neill, A. Rabson, and M. Martin.** 1985. Nucleotide sequence of a full-length human endogenous retroviral segment. J. Virol. **54:**764–772.
16. **Ross, B. M., J. E. Looney, and J. D. Harding.** 1986. Nucleotide sequence of a mouse tRNA Leu gene. Nucleic Acids Res. **13:**5567.
17. **Rowe, W. P., J. B. Humphrey, and F. Lilly.** 1973. A major genetic locus affecting resistance to infection with murine leukemia viruses. J. Exp. Med. **137:**850–853.
18. **Weber, I. T., M. Miller, M. Jaskolski, J. Leis, A. M. Skalka, and A. Wlodawer.** 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. Science **243:**928–931.
19. **Wills, J. W., and R. C. Craven.** 1991. Form, function, and use of retroviral Gag proteins. AIDS **5:**639–654.
20. **Woodcock, S., J. P. Mornon, and B. Henrissat.** 1992. Detection of secondary structure elements in proteins by hydrophobic cluster analysis. Protein Eng. **5:**629–635.
21. **Xiong, Y., and T. Eickbush.** 1990. Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J. **9:**3353–3362.