Scheme for the generation of a truncated endogenous murine leukaemia virus, the \textit{Fv-4} resistance gene

Yutaka Inaguma,† Takamasa Yoshida and Hidetoshi Ikeda*

Laboratory of Experimental Pathology, Aichi Cancer Center Research Institute, 1-1 Kanokoden, Chikusa-ku, Nagoya 464, Japan

The \textit{Fv-4} resistance (\textit{Fv-4} $^r$) gene is a truncated endogenous murine leukaemia virus (MuLV) containing a 3' portion of \textit{pol}, the entire \textit{env} gene and the 3' long terminal repeat. Env expression renders mice resistant to infection by ecotropic MuLVs, probably via receptor interference. Previous studies have suggested that the flanking cellular sequences are also important for \textit{Fv-4 env} gene expression. To establish how the truncated retrovirus was generated and the nature of the cellular sequences involved, the \textit{Fv-4} susceptible (\textit{Fv-4} $^s$) allele DNA was cloned, and its restriction map and nucleotide sequence were compared with those of the \textit{Fv-4} $^r$ allele. A likely mechanism for generation of the truncated endogenous MuLV is suggested by the results; integration of a prototype MuLV provirus at a site within the \textit{Fv-4} $^r$ allele about 6 to 8 kb downstream of a non-retroviral promoter region, followed by deletion of the 5' half of the provirus, with an accompanying loss of only 7 or 10 bp of cellular flanking sequences. The deletion may have led to the expression of the \textit{Fv-4} $^r$ \textit{env} gene under control of the non-retroviral promoter.

Introduction

The mouse \textit{Fv-4} gene determines susceptibility to infection by ecotropic murine leukaemia virus (MuLV) and thereby to tumour induction by the same agent (Kai et al., 1976; Suzuki, 1975). The \textit{Fv-4} locus, located on chromosome 12 (Ikeda et al., 1981; Okada et al., 1981), has two alleles, the resistance (\textit{Fv-4} $^r$) and the susceptible (\textit{Fv-4} $^s$) alleles (Suzuki, 1975). The \textit{Fv-4} $^r$ gene product is an approximately 80K cell surface glycoprotein which is equivalent to the envelope (env) glycoprotein of MuLV (Ikeda & Odaka, 1983, 1984). The \textit{Fv-4} $^s$ gene has been cloned and shown to be a truncated endogenous MuLV containing the entire \textit{env} region (Ikeda et al., 1985; Ikeda & Sugimura, 1989). The \textit{Fv-4} $^s$ env glycoprotein is constitutively expressed on the surface of lymphoid cells and fibroblasts from \textit{Fv-4} $^s$ mice (Ikeda & Odaka, 1984). The resistance function is probably mediated via receptor interference by the \textit{Fv-4} $^s$ env protein. This is supported by the observation that when the cloned \textit{Fv-4} $^s$ DNA was introduced into virus-susceptible cells, the transformed \textit{Fv-4} $^s$ env$^+$ cells became resistant to infection only by ecotropic MuLV strains, not by dualtropic or amphotropic MuLVs (Ikeda & Sugimura, 1989). The ecotropic MuLVs use the same virus receptor of the murine cell, \textit{Rec-1} (Albritton et al., 1989). Consequently, productive infection by an ecotropic MuLV interferes with superinfection of the infected cells by MuLVs of the same ecotropic group (Hartley & Rowe, 1976).

The general structure of endogenous MuLV is long terminal repeat (LTR)–gag–pol–env–LTR, whereas the \textit{Fv-4} gene consists of only Apol–env–LTR, lacking the 5' half of the typical proviral genome (Ikeda et al., 1985; Ikeda & Sugimura, 1989). The \textit{Fv-4} $^r$ gene is thought to have been generated some 500000 years ago in ancestral South-east Asian wild mice, \textit{Mus musculus castaneus} (Inaguma et al., 1991), based on the distribution pattern of the \textit{Fv-4} $^r$ carrier among wild mouse subspecies (Inaguma et al., 1991; Kozak & O'Neill, 1987; Odaka et al., 1978, 1981). The \textit{env} gene of this truncated provirus probably belongs to a MuLV type uniquely present in the \textit{M. m. castaneus} population (Inaguma et al., 1991; Kozak & O'Neill, 1987). Since most endogenous MuLVs are not associated with such a strong \textit{in vivo} resistance to exogenous infection as those carrying the \textit{Fv-4} $^r$ gene, the resistance function ascribed to this gene might be due to its unique structure. Previously it was suggested that the upstream cellular sequence of the truncated \textit{Fv-4} $^r$ MuLV is important for efficient expression of the \textit{env} gene and, therefore, possibly for the resistance function (Ikeda & Sugimura, 1989). Hence, in an attempt to understand the origin of the upstream region and the process leading to generation of the truncated provirus genome, the \textit{Fv-4} $^r$ allele, the putative target DNA region for integration of a prototypic \textit{Fv-4} $^r$ MuLV, was molecularly cloned. Two

† Present address: Institute for Developmental Research, Aichi Prefectural Colony, Kamiya-cho, Kasugai, Aichi 480-03, Japan.

0001-0797 © 1992 SGM
alternative processes were envisioned to explain the generation of the truncated MuLV: a deletion of the 5' half of the prototypic endogenous MuLV or an insertion of a cellular sequence in the middle of the endogenous MuLV. Comparison of the restriction maps and the nucleotide sequences of the Fv-4' and Fv-4" alleles suggested that the truncation resulted predominantly from a viral deletion process. In the Fv-4' allele, a promoter activity was found upstream of the putative integration site.

Methods

DNA cloning and DNA constructions. Cloning of Fv-4' DNAs has been described (Ikeda et al., 1985; Ikeda & Sugimura, 1989). For the cloning of Fv-4' DNAs, the Fv-4' 3' cellular flanking sequence (Fv4'3FL) probe (Ikeda et al., 1985) was used. An EMBL3 phage genomic DNA library of NZB/Jms laboratory mice (Fv-4') was a gift of Drs H. Matsushima and M. Shibuya, Institute of Medical Science, University of Tokyo, Japan. Fv-4'–Fv-4" chimeric DNA constructs were made (see Fig. 3). The R fragment of the Fv-4" allele is a 4.9 kb HindIII–EcoRI fragment derived from pFv4 (Ikeda et al., 1985; Masuda & Yoshikura, 1990). The HindIII site is in the pol region, and is located 437 bp upstream from the splice acceptor site for the env gene. The SAB fragment is a 5.9 kb SaI–HindIII fragment from pFv4Sa. SC from pFv4Sa was derived from a 2.7 kb HindIII–SphI fragment whose ends were further modified to KpnI–HindIII fragments. SAB-R and SC-R constructs were produced by ligating the SAB (SaI–HindIII) and SC (KpnI–HindIII) fragments with the R (HindIII–EcoRI) fragment and insertion into pUC19.

The pOLcat (xSphI) vector was used to analyse promoter activity. The progenitor pOLcat (Yamaguchi et al., 1989) is a pSV2cat-derived vector (Gorman et al., 1982) lacking the simian virus 40 promoter and enhancer and containing a pUC19-derived multiple cloning site located upstream of the CAT gene. The SphI site of pOLcat was deleted in pOLcat (xSph) to remove an initiation codon.

DNA sequencing. Nucleotide sequences were determined by the dideoxynucleotide chain termination method (Sanger et al., 1977) on inserts in pUC18, using an oligonucleotide primer (Takara Shuzou Co.).

DNA transfection into mouse cells. NIH-3T3 cells were transfected with pSV2neo (Southern & Berg, 1982) (1 μg) and plasmids containing the SAB-R or SC-R construct (10 μg) by calcium phosphate precipitation (Okayama & Berg, 1983). Neo' cells were selected in medium containing G418 (400 μg/ml) (Gibco).

Immunofluorescence (IF) staining. Fv-4' env expression on the cell membrane was determined with Fv-4' env-specific alloantiserum (BALB/c anti-BALB/Fv-4') (Ikeda & Odaka, 1984). Transfected NIH-3T3 cells were trypsinized, washed with DMEM containing 1% foetal calf serum, incubated for 30 min at 4 °C with the antiserum, washed, and reacted with fluorescein isothiocyanate-conjugated rabbit anti-mouse immunoglobulin for 30 min at 4 °C. Fluorescent cells were counted with a cryofluorograph (FACSStar, Becton Dickinson).

Chloramphenicol acetyltransferase (CAT) assay. The DEAE–dextran method was used for transfection into NIH-3T3 cells which had been seeded at a density of 4 × 10^5 per 10 cm dish, 12 to 18 h prior to transfection. Cell extracts were prepared at 48 h after transfection. CAT activity was measured by a modification of methods described (Gorman et al., 1982). Conversion rates of 14C-labelled chloramphenicol (Amersham) to acetylated forms were calculated by counting the radioactivities of acetylated and non-acetylated chloramphenicol spots obtained from thin layer plates. Relative CAT activities were further normalized to protein concentrations of each cell extract. Protein determinations were made using a Bio-Rad protein assay kit.

Results

Structure of Fv-4' alleles

The restriction map of the Fv-4' allele is shown in Fig. 1. Using the Fv4 3'FL probe derived from a 3' cellular flanking sequence of the Fv-4' allele (Fig. 1), two Fv-4' DNA clones, pFv4Sa and pFv4Sb, were isolated from a genomic DNA library of the NZB/Jms (Fv-4") laboratory mouse. These clones span about 25 kb of the Fv-4' allele (Fig. 1). No MuLV sequence was detected in these clones by Southern blot hybridization with probes from either the entire Friend MuLV DNA (Oliff et al., 1980) or Fv-4' env DNA (Ikeda et al., 1985) (data not shown).

Gross comparison of the restriction maps indicated that the Fv-4' and Fv-4" DNAs were virtually identical except for the presence of the retroviral sequences in the Fv-4' DNA (Fig. 1). This suggested that the truncation of the endogenous MuLV could have resulted from a deletion of the putative, prototypic endogenous MuLV, rather than from an insertion of the cellular sequence from a distinct chromosomal region into the middle of the integrated provirus.

Possible mechanism for generation of the Fv-4' gene

Comparison of the restriction maps suggested that the deletion included a short adjacent 5' cellular sequence, probably no more than 1 kb in size (Fig. 1). To define the deletion length more precisely, a short region (312 bp) of the Fv-4' DNA where the retroviral sequence is located in the Fv-4' DNA was sequenced from an EcoRI site at map position 12 kb toward map position 13 kb (Fig. 1). The first 236 bp of the Fv-4' sequence matched perfectly the Fv-4' cellular sequences (until base number 36 of Fig. 4 in the paper of Ikeda et al. (1985)) which flanks the 5' end of the truncated pol region, and after a 3 bp gap (CTT), the remaining 73 bp also matched the cellular sequences (from base number 647 of Fig. 5 in the same paper) at the 3' terminus of the LTR (Fig. 2a and c). The mismatched nucleotides of Fv-4' comprised 3393 bp of retroviral sequences (Fig. 2c).

Based on these results, a most probable scheme outlining the process of Fv-4' gene generation is presented (Fig. 2). Integration of MuLV proviral DNA into host cell DNA typically generates a 4 bp target sequence duplication at both ends of the endogenous
Identification of promoter activity upstream of the integration site in the Fv-4' allele

The possibility that the prototypic Fv-4' provirus integrated downstream of a promoter region of a putative gene in the Fv-4' allele was tested by using two different reporter genes, the Fv-4' MuLV gene and the CAT gene.

The Fv-4' MuLV (R in Fig. 3a) was ligated to two Fv-4' fragments, SAB and SC, which are located 8 kb to 3 kb and 3 kb to +0-3 kb from the integration site, respectively. These DNA constructs (SAB-R and SC-R) and a neomycin resistance gene (pSV2neo) were introduced into NIH-3T3 cells. After selection with G418, the resultant neo' colonies were pooled and analysed for Fv-4' env expression by membrane IF staining using antisera specific to the Fv-4' env glycoprotein (Ikeda & Odaka, 1984). Env expression was strong in half of the SAB-R-transfected cells but equivalently faint expression was seen in both the SC-R-transfected cells and the parental NIH-3T3 cells (Fig. 3b). These observations were consistently reproducible.

RNAs extracted from these transfected cells were analysed by Northern blotting using an Fv-4' env-specific probe. No Fv-4' env RNA was detected in the SC-R transfecteds (Fig. 3c, lane 5), whereas two RNAs of about 3-0 kb and 3-4 kb were detected in the SAB-R transfecteds (lanes 4 and 11). The 3-0 kb RNA was identical in size to the original Fv-4' env RNA from the Fv-4' mice (lanes 9, 10 and 11), but the origin of the 3-4 kb RNA was unknown. The two RNA species in the pooled cells did not arise from a mixture of cells expressing either of these RNAs, since all of the 10 Fv-4' env' cell clones that were selected by limiting dilution from the pooled SAB-R-transfected cells expressed both sizes of transcripts (lanes 6, 7 and 8). In nine of the 10 cell clones, the larger RNA was more intense than the shorter RNA (lanes 6 and 8). The remaining clone expressed the same two RNAs with identical intensity (lane 7).

As reported, when the Fv-4' DNA (λ Fv4-17 in Fig. 1) containing the Fv-4' upstream region was introduced into the same fibroblast cell line, only the 3-0 kb Fv-4' env RNA was detected in the transfectants (Ikeda & Sugimura, 1989). Thus, the generation of the larger RNA in the SAB-R transfectants may be due to either the artificial DNA construct or to possible sequence differences between the Fv-4' and Fv-4' alleles.

By using the transient CAT gene expression system, the promoter region was further localized. The SAB region was separated into SA and SB regions which were located 8 to 6 kb and 6 to 3 kb upstream from the integration site (Fig. 4). These fragments were placed in front of the CAT gene of the pOLcat-derived vector (see Methods), which has neither an enhancer nor a promoter.

Mutant MuLV (Dhar et al., 1980; Craigie et al., 1990). Assuming that the truncated Fv-4' MuLV was originally a normal endogenous MuLV, the target sequence would be CTTT because these four bases occur at the 3' end of the LTR (Fig. 2c). The 5' end of the prototypic endogenous MuLV should also have been flanked by CTTT (Fig. 2b).

Upstream of the 3 bp mismatched sequence (CTT), the first matched sequence ended in TGG (Fig. 2a). It was previously anticipated that this TGGG in the Fv-4' DNA was derived from the prototype MuLV (Ikeda et al., 1985) because of its identity with sequences of ecotropic AKV (Herr et al., 1982) and ecotropic Moloney (Shinnick et al., 1981) MuLVs. However, since the same TGG existed at the same position in the Fv-4' allele (Fig. 2a and c), it is uncertain whether the deletion involved 10 bp (deletion 1 in Fig. 2b) or 7 bp (deletion 2) of the adjacent cellular sequence. The deletion might have been mediated by the 3 bp (TGG) homology between the cellular sequence and the retroviral sequence.

Other mechanisms can also be considered (see Discussion).
Fig. 3. Detection of a transcription promoter activity in the \( Fv-4' \) allele. Chimeric DNA constructs of the \( Fv-4' \) and \( Fv-4' \) alleles were analysed for their ability to induce both \( Fv-4' \) env RNA and antigens on the cell surface. (a) Origins of DNA fragments used for the DNA constructs. The boxes in the \( Fv-4' \) allele indicate the retroviral sequences. The triangle on the \( Fv-4' \) allele indicates the provirus integration site. The restriction enzyme sites shown are BamHI (B), EcoRI (E), HindIII (H), KpnI (K), Sall (S) and SpII (Sp). (b) Expression of \( Fv-4' \) env antigen on NIH-3T3 cells: (i) control; (ii) NIH-3T3 cells transfected with SAB-R DNA; (iii) NIH-3T3 cells transfected with SC-R DNA. The SAB-R or SC-R DNA was cotransfected with pSV2neo into NIH-3T3 cells. Neo' colonies were selected in medium containing 400 \( \mu \)g/ml G418. Twenty-seven Neo' clones of SAB-R-transfected cells or about 200 Neo' clones of SC-R-transfected cells were mixed, passaged twice, and analysed for \( Fv-4' \) env expression on the cell surface by IF flow cytometry. (c) \( Fv-4' \) env RNA expression in NIH-3T3 cells transfected with the SAB-R construct as detected by Northern blot analysis. Lane 1, BALB/c mouse (\( Fv-4'' \)) liver (5 \( \mu \)g RNA). Lanes 2 and 9, BALB/c-\( Fv-4' \) congenic mouse liver (5 \( \mu \)g RNA). Lane 3, NIH-3T3 cells (5 \( \mu \)g RNA). Lane 4, NIH-3T3 cells transfected with SAB-R DNA, a mixture of 27 Neo' colonies (5 \( \mu \)g RNA). Lane 5, NIH-3T3 cells transfected with SC-R DNA, a mixture of about 200 Neo' colonies (5 \( \mu \)g RNA). Lanes 6 to 8, SAB-R-transfected cell clones (5 \( \mu \)g RNA). Lane 10, a mixture of BALB/c-\( Fv-4' \) liver (5 \( \mu \)g RNA) and SAB-R-transfected NIH-3T3 cells (0.5 \( \mu \)g RNA). Lane 11, SAB-R-transfected NIH-3T3 cells (0.5 \( \mu \)g RNA). Autoradiograph exposure times for lanes 6 to 8 were one-tenth those for the remaining lanes.
Promoter activity was detected in the SA region (Fig. 4, lane 2) but not in the SB (lane 3) or SC regions (lane 4). As the SA fragment is shorter than the SB and SC fragments, it was possible that the failure to detect promoter activity in the SB and SC regions was due to the greater distances between possible promoters and the CAT gene. However, this possibility seemed unlikely because even the SAB fragment was able to induce detectable levels of CAT activity (Fig. 4, lane 1).

Discussion

The present study characterized DNA clones of the \(Fv-4^r\) allele in comparison with those of the \(Fv-4^s\) allele. The restriction maps and nucleotide sequences suggested that the truncated endogenous MuLV was generated by deletion of a complete endogenous virus or by integration of a truncated provirus, but not by an integration of a large (more than 12 kb) non-retroviral DNA fragment into a complete endogenous MuLV. The most probable process is presented in Fig. 2, based on the deletion model of a complete endogenous MuLV. The deletion appeared to involve about 5 kb of the 5' half of the endogenous MuLV genome and only 7 or 10 bp of the cellular flanking sequence, so that the deletion seems to have preserved most of the upstream cellular sequences. As a result of the deletion, the endogenous MuLV lost the 5' LTR that generally controls expression of the endogenous MuLV genome. Functional analyses of the cloned \(Fv-4^r\) DNA demonstrated the existence of promoter activity in the region upstream from the integration site. Thus, this DNA rearrangement may have allowed \(Fv-4^r\)\(\text{env}\) expression to be controlled by a non-retroviral promoter which was present prior to integration. The rearrangement may also be crucial for the unique \textit{in vivo} resistance to ecotropic MuLV infection.

It is also possible that a truncated provirus integrated into the \(Fv-4^s\) allele. In this case, the 3' end of the truncated provirus would integrate normally and the 5' end by some illegitimate mechanism possibly involving the TGG sequence (Fig. 3). However, as suggested by a recent study using the \textit{in vitro} integration system that proviral integration requires intact ends of proviral DNA (Craigie \textit{et al.}, 1990), it is not clear whether the truncated \(Fv-4^r\) provirus could actually be integrated by such a mechanism.

Promoter activity found in the \(Fv-4^s\) allele was located about 3 kb to 8 kb upstream from the provirus integration site. The promoter can induce the expression of both the \(Fv-4^r\)\(\text{env}\) gene in the stable expression system and the CAT gene in the transient expression system. The promoter may also be that used for the \textit{in vivo} expression in \(Fv-4^r\) mice, since one of the two RNAs expressed by the \(Fv-4^r\)-\(Fv-4^s\) MuLV chimeric DNA construct was identical in size to the original \(Fv-4^r\)\(\text{env}\) RNA (Fig. 3). The R fragment contains the splice acceptor site of the \textit{env} region, the entire \textit{env} region and the poly(A) signal in the 3' LTR, spanning about 2-7 kb.
Since the two env RNAs in the SAB-R transfecants were about 3.0 kb and 3.4 kb in size and hybridized with an SA probe (data not shown), the SAB region may contain a promoter and short exon(s).

Although a promoter activity was identified in the Fv-4 allele, the nature of this allele is obscure. The Fv-4 env antigen is expressed in a variety of organs (Ikeda & Odaka, 1984); therefore it was considered that, if the promoter regions are the same for these alleles, then the putative gene in the Fv-4 allele may also have been expressed in various organs. However, preliminary Northern blot hybridization analysis using several probes cut from the Fv-4 cellular sequences failed to detect any RNA specific for the Fv-4 allele (data not shown). As a possible explanation, the high env expression from the Fv-4 allele could have been an effect of enhancer sequences in the 3' LTR on the upstream promoter, although Fv-4 promoter region can function without an LTR in fibroblasts as demonstrated in the transient CAT expression system (Fig. 4). In addition, the promoter and its adjacent exon regions may not necessarily be the same for the two alleles, as implied by the unexpectedly longer RNA expressed in the SAB-R transfecants (Fig. 3). Alternatively, RNA expressed from a possible gene in the Fv-4 allele might be unstable or too low in abundance to be detected by standard Northern blotting. It should be pointed out that these alleles appeared to have diverged about 5 x 10^5 years ago (Inaguma et al., 1991), so that mutations in these alleles may have occurred during evolution. The presence of the truncated endogenous MuLV in this locus does not seem to affect normal genetic functions of mice, because the BALB/c-Fv-4 congenic mice, homozygous for the Fv-4 allele, appear to have a lifespan similar to that of BALB/c (Fv-4) mice.

We thank M. Hirose and M. Shibuya for technical assistance, J. K. Makino for reviewing the manuscript, H. Matsushima and M. Shibuya for providing the recombinant phage library of Fv-4 mouse DNA, and K. Kojie for cytfluorographic analysis. This work was supported in part by a grant-in-aid from the Ministry of Education, Science and Culture of Japan.

References


Gorman, C. M., Moffat, L. F. & Howard, B. H. (1982). Recombinant genomes which express chloramphenicol acetyltrans-

ferase in mammalian cells. Molecular and Cellular Biology 2, 1044–1051.


(Received 18 December 1991; Accepted 8 August 1992)