JOURNAL OF VIROLOGY, Sept. 1984, p. 695–705
0022-538X/84/090695-11$02.00/0

Vol. 51, No. 3

Molecular Cloning of Biologically Active Rauscher Spleen Focus-Forming Virus and the Sequences of Its *env* Gene and Long Terminal Repeat

RICHARD K. BESTWICK,* BRUCE A. BOSWELL, AND DAVID KABAT

Department of Biochemistry, School of Medicine, Oregon Health Sciences University, Portland, Oregon 97201

Received 9 April 1984/Accepted 1 June 1984

Rauscher and Friend spleen focus-forming viruses (R- and F-SFFVs) cause similar progressive erythroleukemias dependent upon a virus-encoded membrane glycoprotein. Moreover, these SFFV glycoproteins are immunologically related to each other and to the recombinant-type glycoproteins encoded by the *env* genes of dual tropic murine leukemia viruses. To better understand these diseases and the viral origins, we isolated a pathogenically active molecular clone of R-SFFV proviral DNA, sequenced its 3′-terminal 2,163-base-pair (bp) region, and compared these sequences with previously determined sequences of F-SFFV. The 516-bp R-SFFV long terminal repeat is highly homologous to those of F-SFFV and Friend murine leukemia virus, although only the latter contains a 65-bp direct repeat in its U3 region. The *env* gene of R-SFFV encodes a glycoprotein with 408 amino acids that is identical in its basic domain organization to the glycoprotein of F-SFFV. Thus, the junctions between the dual tropic-related and ecotropic sequences occur at the same nucleotide, and both SFFV *env* genes contain identical 585-bp deletions in their ecotropic domains and single-bp insertions which cause premature terminations at the same amino acid in their ecotropic p15E domains. Consistent with their independent origins, however, the *env* sequences of R- and F-SFFV are distinctive in both their 5′ dual tropic-related and 3′ ecotropic-related domains. Furthermore, there are several consistent amino acid differences between the polycythemic F-SFFV sequences and the anemia-inducing R-SFFV sequence. The striking similarities of the independently formed F- and R-SFFV *env* genes imply that all of the glycoprotein domains arranged in a precise organization may be required for its leukemogenic activity.

The Rauscher and Friend erythroleukemia viruses contain two components. One component, the spleen focus-forming virus (SFFV), is replication defective and causes foci of proliferating erythroid cells in spleens of infected mice (3, 13, 48). The second component is a replication-competent murine leukemia virus (MuLV) that serves as a helper virus for SFFV. The diseases induced by these viral complexes develop in two phases, an initial hyperplasia of erythroblasts resulting in splenomegaly within 1 to 2 weeks, followed later by the emergence of erythroleukemia cells that can be passaged indefinitely in secondary recipients or in culture (31, 54). Although the original isolates of Rauscher and Friend SFFV (R- and F-SFFV) cause anemia, apparently because of ineffective erythropoiesis by infected erythroblasts, F-SFFV variants were found to cause polycythemia (2, 36, 49). With these F-SFFVs, the infected erythroblasts differentiate to some extent, even in the absence of substantial erythropoietin (19, 40a, 49).

The R- and F-SFFVs are structurally closely related to the dual tropic *env* gene recombinant class of replication-competent MuLVs (dtMuLVs) (1, 4, 5, 8, 10, 40, 51, 52, 55). Moreover, the latter viruses have also been implicated as causes of a variety of progressive hematopoietic leukemias and lymphomas (9, 12, 17, 49). Based on these structural and pathogenic similarities, it has been proposed that dtMuLVs and SFFVs might cause proliferation of hematopoietic stem cells by related mechanisms (4, 42). Unlike dtMuLVs, however, the F- and R-SFFV *env* genes encode smaller glycoproteins with apparent M_rs of 55,000 (gp55) and 54,000 (gp54), respectively (10, 42), and these glycoproteins are processed relatively inefficiently from the rough endoplasmic reticulum into the plasma membranes; they are also excluded from budding virions (4, 10, 40, 40a, 42). Recently, it has been demonstrated that SFFV pathogenesis is eliminated by mutations within the gp54-55-encoding *env* genes of F-SFFV (27, 41) and R-SFFV (30) and that intragenic suppressor mutations within *env* can cause restoration of leukemogenic activity (C. Machida and D. Kabat, submitted for publication). These results have established that the recombinant-type *env* genes of SFFVs are oncogenes which cause progressive erythroleukemias.

Despite their similarities, however, the F- and R-SFFVs are also structurally distinct. For example, F-SFFV contains several deletions which cause premature termination of its *gag* polyprotein and inactivation of its *pol* gene (5, 8, 11, 55). On the contrary, R-SFFV encodes an intact *gag* polyprotein which is functional by the criterion that it organizes into a virus-like particle that spontaneously buds from cells in the absence of a helper MuLV (4). Furthermore, these released particles contain enzymatically active reverse transcriptase, suggesting that the R-SFFV *pol* gene is also functional (4). These results, combined with nucleic acid analyses of R-SFFV proviral DNA and mRNAs, have suggested that R-SFFV may have derived from a dtMuLV by changes within the *env* gene, which include a partial deletion of 0.6 to 0.7 kilobases (kb) (4).

The minimal difference between the genomes of R-SFFV and the dtMuLVs presumably must account for the enhanced leukemogenic potential of R-SFFV. Furthermore, because the Friend (14) and Rauscher (38) viral complexes are of independent origin, a comparison of their SFFV-encoded *env* gene sequences could highlight critical structural features in SFFV pathogenesis. For these reasons and to compare in detail a wild-type R-SFFV with the non-

* Corresponding author.

leukemogenic R-SFFV mutants that have recently been isolated in this laboratory (30), we molecularly cloned the circular unintegrated proviral forms of R-SFFV. We report here the restriction endonuclease map, the long terminal repeat (LTR) and *env* gene sequences, and the infectivity of one of these clones.

MATERIALS AND METHODS

Cells and viruses. All cell lines were grown in Dulbecco modified Eagle medium (GIBCO Diagnostics, Madison, Wis.) supplemented with fetal calf serum and antibiotics. The Rauscher virus-infected NIH/3T3 fibroblast line (Rauscher virus NIH clone 18) releases R-SFFV and R-MuLV in a ratio of ca. 10:1. NIH/3T3 fibroblasts were used as the uninfected cell line for preparation of R-SFFV viral DNA intermediates and for DNA transfection experiments.

Preparation of unintegrated closed circular DNA. Plastic roller bottles (490 cm^2; Corning) were seeded with 2×10^7 NIH/3T3 cells and 0.5×10^7 Rauscher virus NIH clone 18 cells, both of which were in an actively growing state. Infection by cocultivation (44) was allowed to proceed for 24 to 30 h in the presence of 8 μg of Polybrene per ml. Low-molecular-weight extrachromosomal DNA was isolated by the procedure of Hirt (20) and was electrophoresed in a 0.8% low-melting agarose gel. DNA was purified, using a method described by Maniatis et al. (32), from that region of the gel corresponding to covalently closed circular proviral DNA containing both single and double LTR molecules.

Cloning. One microgram of the size-fractionated Hirt supernatant DNA was treated with *Eco*RI, extracted with phenol-chloroform-isoamyl alchohol (25:24:1), ethanol precipitated, and ligated to 5 μg of λ Charon 3A *Eco*RI arms. All ligations were done with T4 DNA ligase from New England Nuclear Corp. or Collaborative Research, Inc. In vitro packaging was carried out according to Maniatis et al. (32), using protocol II. Approximately 5×10^4 plaques were screened with a ^{32}P-labeled SFFV-specific hybridization probe described previously (4, 7). This probe is the 620-base *Bam*HI-*Eco*RI *env* gene fragment from molecularly cloned F-SFFV (28). Positive plaques were plaque purified and grown in liquid culture to obtain DNA for analysis and subcloning. The DNA from one of the positive λ clones was digested with *Eco*RI, ligated to *Eco*RI-cut pBR322, and transformed into HB101. Restriction map analysis, DNA sequencing, and infectivity experiments were done with this pBR322 subclone.

Restriction endonuclease analysis. Hirt supernatant, λ, or plasmid DNAs were digested with restriction endonucleases according to the recommendations of the supplier (New England Biolabs or Bethesda Research Laboratories). Restricted DNA was fractionated on agarose gels containing 0.5 μg of ethidium bromide per ml, photographed under UV light, and, when necessary, transferred to diazophenylthioether paper by the methods of Seed (45), Southern (46), and Wahl et al. (53). After transfer, the diazophenylthioether paper was prehybridized for at least 1 h at 42°C in 50% formamide–5× SSPE (1× SSPE is 0.18 M NaCl, 10 mM NaPO4 (pH 7.7), 1 mM EDTA)–5× Denhardt (0.1% polyvinylpyrrolidone, 0.1% Ficoll 400, 0.1% bovine serum albumin)–sonicated denatured salmon sperm DNA (200 μg/ml)–1.0% glycine–0.1% sodium dodecyl sulfate (SDS). Hybridization was carried out in the same buffer containing 10% dextran sulfate and ^{32}P-labeled DNA at 10^4 to 10^6 cpm per ml. Hybridization was allowed to proceed overnight at 42°C, after which the filters were washed two times for 5 min at room temperature in 2× SSPE–0.05% SDS followed by a single 30-min 50°C wash in 0.1× SSPE–0.05% SDS. The washed filters were exposed to Kodak XAR film at −70°C for variable lengths of time, depending on the signal intensity. Diazophenylthioether filters were prepared for reuse by soaking in 0.5 M NaOH for 5 to 10 min, followed immediately by two 5-min washes in 2× SSPE–0.05% SDS. Prehybridization and hybridization were then performed as described above.

Radiolabeled DNA hybridization probes were prepared by either of two methods. Nick translation of intact plasmid DNA was performed as described by Maniatis et al. (32), using [α-^{32}P]dATP or dCTP (New England Nuclear) at 800 Ci/mmol. Isolated restriction fragments were labeled by T4 DNA polymerase (New England Nuclear)-catalyzed replacement synthesis by the method of O'Farrell et al. (35), using the same radiolabeled nucleotides described above. Specific activities of labeled DNAs by either of these methods exceeded 10^8 dpm per μg of DNA.

Transfection and infectivity assays. NIH/3T3 cells were cotransfected with an R-SFFV molecular clone that was constructed in vitro such that the genome was not permuted at the *Eco*RI site (see below) and pSV2neo (47). Transfections were performed by the method of Graham and Van der Eb (16) with the modifications of Parker and Stark (35). Cotransfectants were selected by growth in media containing 275 μg of G418 (GIBCO) per ml. Drug-resistant colonies were screened for expression of R-SFFV-encoded gp54 by L-[^{35}S]methionine labeling, radioimmune precipitation with anti-gp70 antisera, and SDS-polyacrylamide gel electrophoresis as described previously (42). A colony expressing gp54 was superinfected with R-MuLV, and 0.5 ml of the virus-containing media from these cells was used for tail vein injection of 4- to 6-week-old female Swiss mice. Gross pathologic analyses of spleens, measurements of hematocrits, and preparation of spleen homogenates were done by previously described methods (30). Spleen sections were also stained with hemotoxylin-eosin for histological examinations.

Nucleotide sequence analysis. Large restriction fragments of pBC-6 and pBC-10 (see below) encompassing the envelope gene and the LTR were isolated with low-melting agarose gel electrophoresis as described by Maniatis et al. (32). These DNA fragments were cleaved with one or a combination of the following restriction endonucleases: *Alu*I, *Msp*I, *Taq*I, *Sau*3A, and *Hin*PI. The resulting restriction fragments were then ligated into the appropriate cloning sites of the M13 vectors, MP8 and MP9. Clones of specific fragments were isolated by forced cloning or screening of clone banks with spot blots and M13 hybridization probes. The specific protocols for these methods are described by Messing (33). Sequencing was performed by the Sanger dideoxynucleotide termination method (43), using the protocol of Messing (33). The sequencing primers were either the 15mer or 17mer from New England Biolabs or the 15mer from New England Nuclear. The Klenow fragment of *Escherichia coli* DNA polymerase I was obtained from New England Nuclear.

Compilation and analysis of the sequence data was performed on an Apple II Plus microcomputer, using the programs of Larson and Messing (25) and Fristensky et al. (15).

RESULTS AND DISCUSSION

Molecular cloning and restriction endonuclease analysis of unintegrated R-SFFV proviral DNA. The strategy used for cloning the unintegrated circular form of R-SFFV is shown

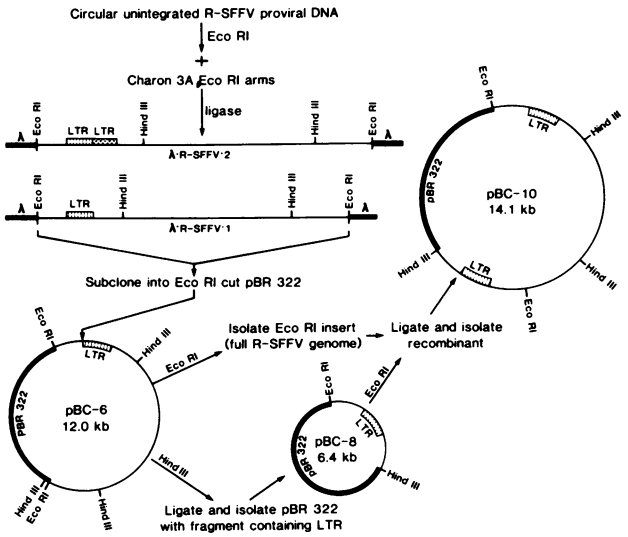

FIG. 1. Strategy used for the construction of R-SFFV molecular clones and a subclone containing a nonpermuted genome that was subsequently used to test for infectivity. Details of the procedure are described in the text.

in Fig. 1. We chose to clone circular proviral DNA because it contained a single *Eco*RI site, which allowed us to clone a full-length copy. As shown in the left panel of Fig. 2, uncut Hirt supernatant proviral DNA (lane 1) showed a major band at 8.3 kb, which corresponds to the linear form with two LTRs and two minor bands (barely visible in this exposure) just below the 4.4-kb marker that represent supercoiled provirus with one or two LTRs. After cleavage with *Eco*RI (left panel, lane 2), the minor bands are seen at 8.3 and 7.7 kb, and the major band is reduced to 7.0 kb. These results indicated that a single *Eco*RI site was located 1.3 kb from one end of the linear proviral DNA. Circular proviral DNA was isolated by preparative agarose gel electrophoresis, cleaved with *Eco*RI, and ligated to the arms of λ Charon 3A. Four recombinant λ · R-SFFV clones were isolated, three with inserts of 8.3 kb and one with an insert of 7.7 kb. Restriction endonuclease analysis of two of these clones compared with a similar analysis of Hirt supernatant DNA is shown in the remainder of Fig. 2. All of the restriction sites determined to be present in the Hirt supernatant proviral DNA were also present in the recombinant λ clones. Additional mapping (data not shown) indicated that the size difference between the two sets of clones was, as expected, due to the presence of either one (7.7 kb) or two (8.3 kb) LTRs as shown in Fig. 1.

The λ · R-SFFV clone containing a single LTR was subcloned into the *Eco*RI site of pBR322 and given the designation of pBC-6 (Fig. 1). pBC-6 was then extensively mapped (Fig. 3) and compared with an equally extensive restriction endonuclease map of R-SFFV Hirt supernatant proviral DNA (data not shown). The maps obtained from both analyses were coincident. Shown in Fig. 4 is a comparison of restriction endonuclease cleavage sites on the restriction maps of R-MuLV (18) and R-SFFV DNAs. The results indicate that the 5′ two-thirds of R-MuLV and R-SFFV are very similar but that substantial differences exist in the 3′ one-third of the genomes, which are known to encompass the envelope glycoprotein genes of both viruses. This substantiates and extends previous results (4) which suggest that R-SFFV differs from its R-MuLV ecotropic helper virus by a substitution of dt related *env* gene sequences and a small deletion (approximately 0.6 kb) in the R-SFFV *env* gene region.

Infectivity. Circularly permuted clones of SFFV require recombination with cotransfected cloned helper viral DNA or with other SFFV DNA molecules to produce infectious SFFV virions (28, 29). To avoid the uncertain nature of helper-SFFV or SFFV-SFFV recombinants, a nonpermuted R-SFFV clone was constructed in vitro. Figure 1 shows a schematic representation of how pBC-6 was manipulated to produce a clone (pBC-10) that contains a contiguous genome. pBC-6 was cleaved with *Hin*dIII and religated to produce a clone (pBC-8) that had the LTR and flanking partial *env* and *gag* gene sequences. pBC-8 was subsequently cleaved with *Eco*RI and ligated to the 7.7-kb *Eco*RI insert (the full R-SFFV genome) from pBC-6. The resulting recombinant (pBC-10) was mapped with restriction endonucleases to ensure the proper orientation of the two viral fragments (data not shown). As shown below, pBC-10 needs only to be transfected into NIH/3T3 cells and rescued with a helper virus to obtain infectious R-SFFV virions.

NIH/3T3 cells were cotransfected with pBC-10 and pSV2-neo, and G-418-resistant colonies were isolated after at least 1 week in selective media. Ten drug-resistant clones were screened for the expression of gp54, and one clone (designated RT cl 2) was positive (Fig. 5). The R-SFFV in RT cl 2 was rescued by superinfection with R-MuLV, and the media from the superinfected cells was used to infect 4- to 6-week-old Swiss mice. Table 1 shows the gross splenic pathologies

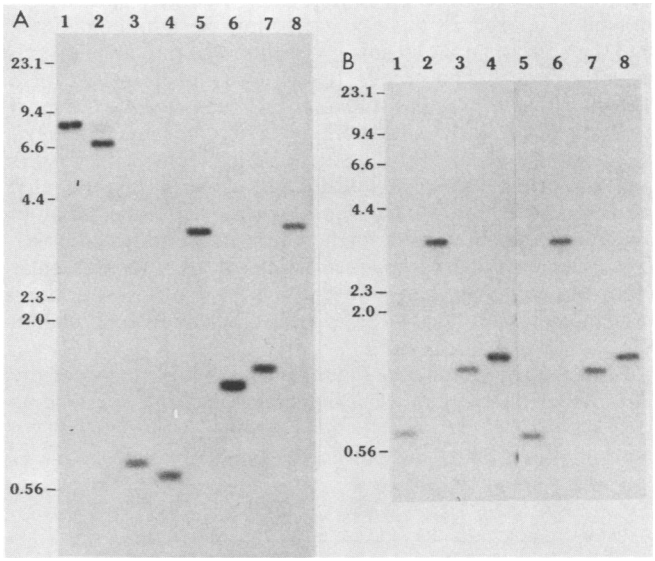

FIG. 2. Restriction endonuclease analysis of R-SFFV unintegrated proviral DNA and two λ · R-SFFV molecular clones. (A) Unintegrated proviral DNA (Hirt supernatant) was treated with various restriction endonucleases, fractionated by agarose gel electrophoresis (0.8%), and subjected to Southern blotting analysis with the SFFV-specific hybridization probe (see text). The enzymes used were as follows: undigested control (lane 1), *Eco*RI (lane 2), *Eco*RI-*Xba*I (lane 3), *Eco*RI-*Bam*HI (lane 4), *Eco*RI-*Kpn*I (lane 5), *Eco*RI-*Hin*dIII (lane 6), *Eco*RI-*Hin*dII (lane 7), and *Eco*RI-*Bcl*II (lane 8). (B) λ · R-SFFV-cloned DNAs containing either one LTR (lanes 1 to 4) or two LTRs (lanes 5 to 8) were digested, fractionated, and blotted as above. The enzymes used were *Eco*RI-*Xba*I (lanes 1 and 5), *Eco*RI-*Kpn*I (lanes 2 and 6), *Eco*RI-*Hin*dIII (lanes 3 and 7), and *Eco*RI-*Hin*dII (lanes 4 and 8). The migration of coelectrophoresed *Hin*dIII-digested λ DNA is indicated in kb along the left margin.

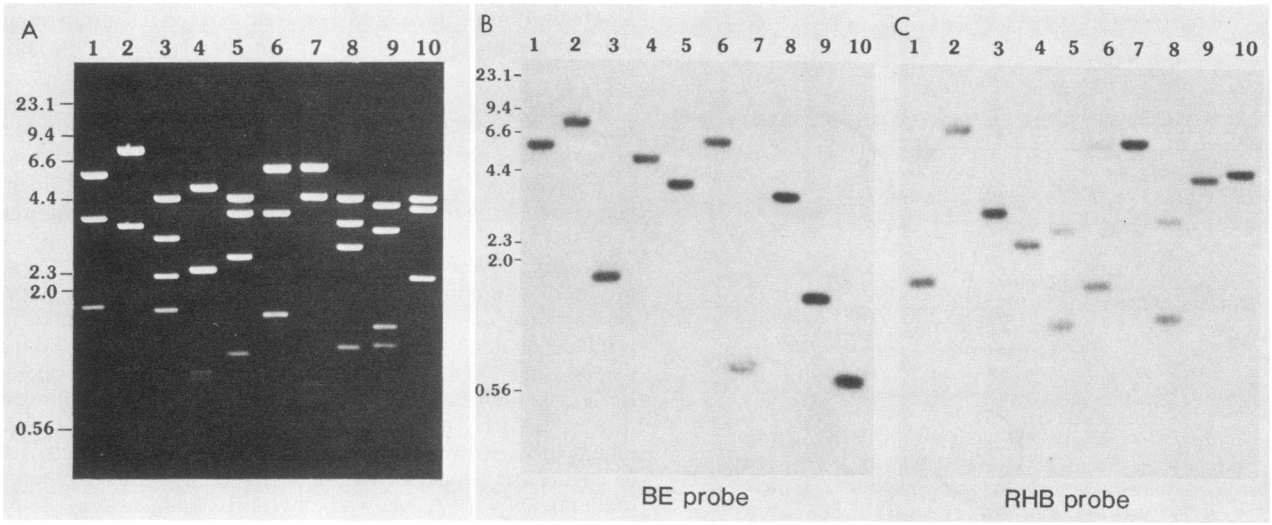

FIG. 3. Detailed restriction endonuclease mapping of a plasmid subclone of R-SFFV. The plasmid subclone pBC-6 (see Fig. 2) was digested with *Eco*RI and then with various other restriction endonucleases, fractionated by agarose gel electrophoreses (0.8%), Southern blotted to diazophenylthioether paper, and hybridized to subgenomic hybridization probes from various regions of the R-SFFV genome. After autoradiography, the signal was removed by soaking the blot in 0.5 N NaOH for 10 min. (A) Photograph of the ethidium bromide-stained gel. (B) Autoradiograph after hybridization to the 0.6-kb *Bam*HI-*Eco*RI *env* gene probe or (C) the 2.1-kb *Eco*RI-*Hin*dIII fragment containing the LTR. Lanes 1 through 10: *Pst*I, *Nru*I, *Bgl*II, *Bgl*I, *Sst*I, *Pvu*I, *Xba*I, *Kpn*I, *Hin*dII, and *Bam*HI. Autoradiography was for 1 h at room temperature. The position of coelectrophoresed *Hin*dIII-digested λ DNA is indicated in kb along the left margin.

which resulted from one set of injections. To amplify the R-SFFV titer by serial passaging (30), enlarged spleens removed at 21 and 28 days postinjection were homogenized, and the cell-free homogenates were injected into an additional set of mice. The results from these injections are also shown in Table 1. Rapid splenomegaly was apparent in most of these mice, and histological analysis confirmed the diagnosis of splenic erythroblastosis. Consistent with results with our original stock of wild-type Rauscher virus, the mice analyzed in Table 1 also reproducibly developed a mild anemia as indicated by their hematocrits (data not shown). By these criteria, pathogenesis caused by the molecular clone of R-SFFV is very similar to our previous descriptions of pathogenesis by the wild-type Rauscher viral complex (30, 42).

DNA sequence analysis. Figure 6 shows that region of the R-SFFV genome that was sequenced and the sequencing

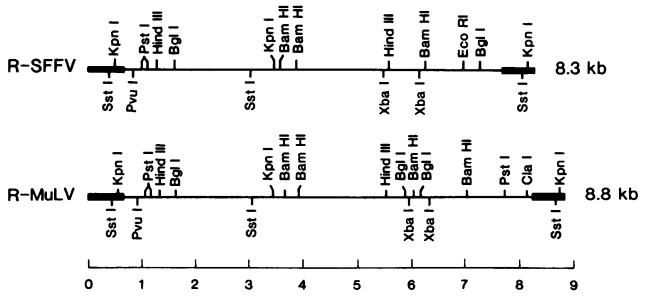

FIG. 4. Restriction endonuclease map of the R-SFFV genome and a comparison of common restriction endonuclease sites in R-MuLV (18). Map units are in kb pairs. Restriction endonucleases that do not cleave R-SFFV are *Xho*I, *Sal*I, *Cla*I, and *Nru*I. In addition to the sites shown for R-SFFV, there are also *Hin*dII sites at 3.5 and 5.4 map units.

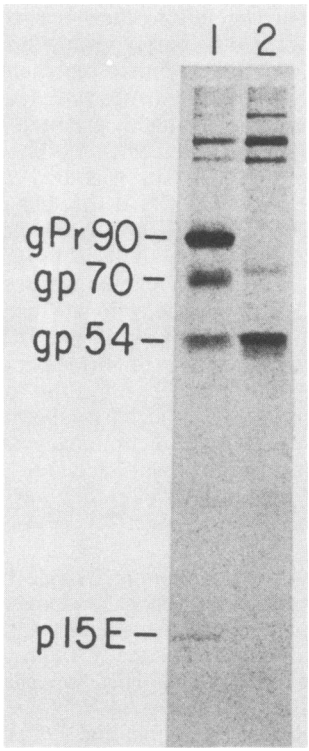

FIG. 5. Electrophoretic analysis of [^{35}S]methionine-labeled proteins from Rauscher virus-infected NIH/3T3 cells (lane 1) and from NIH/3T3 cells stably transfected with molecularly cloned R-SFFV (lane 2). Cells were pulse-labeled with [^{35}S]methionine for 2 h and lysed; the *env*-related proteins were immunoprecipitated with monospecific serum to gp70 *env*. Immunoprecipitated proteins were electrophoresed through 10 to 20% polyacrylamide gels containing 0.1% SDS.

TABLE 1. Virus produced from a molecular clone of R-SFFV causes Rauscher disease

Virus injected[a]	Days after injection	Spleen wt[b] (mg)	Foci[c]
Medium from pBC-10-transfected NIH/3T3 cells[a]	14	225,265	+,+
	21	134,306,307 (231,256)[d]	+,−,−
	28	(386,831,723)[d] 195,219,238	+,+,+
	35	179,191,732	+,+,+
21-Day spleen homogenates[c]	14	380,490	+,+
	21	740,455	+,+
	35	568	+
28-Day spleen	14	550,1655	+,+
	21	1565,1470	+,+
	35	611,446	+,+

[a] A molecular clone of R-SFFV (pBC-10) was transfected into NIH/3T3 cells, and the virus released after superinfection with a helper R-MuLV was injected in 0.5-ml portions into 4- to 6-week-old Swiss mice. Control mice were either injected with R-MuLV alone or with culture media alone (see footnote *b*).

[b] Spleen weights from control mice injected with R-MuLV helper or culture media alone ranged from 100 to 175 mg during the same time course, consistent with previous studies (30).

[c] Foci were seen as pale discolorations of various sizes on the surfaces of spleens fixed in Bouin fixative (3). The two spleens indicated as negative were enlarged and had slight discolorations which were broadly distributed on their surfaces. Evidence for focal splenic erythroblastosis was confirmed in some cases by histological examination.

[d] Those spleens whose weights are in parentheses were homogenized, and the cell-free supernatants were injected into mice. The results from these injections are shown in the bottom two groups.

strategy used. All regions were sequenced at least twice, and for most regions both strands were sequenced. The restriction endonuclease sites used for cloning into M13 were all sequenced except for the *Bam*HI site at nucleotide 227. The nucleotide sequence of the R-SFFV *env* gene and the LTR is shown in Fig. 7. Also shown is the deduced amino acid sequence for the longest open reading frame in this region (gp54), which is 408 amino acids long. There are five potential Asn-X-Thr/Ser glycosylation sites in gp54, and these are denoted in Fig. 7.

The R-SFFV LTR is 516 nucleotides long (Fig. 8) and is very similar to the sequences obtained for other murine retroviruses. It contains the transcriptional regulatory sequences and short terminal repeats that are characteristic of retroviral LTRs, but it lacks the 65- to 75-base-pair repeat in the U3 region which occurs in Moloney MuLV, Moloney murine sarcoma virus, and F-MuLV (22). A comparison of the homology between the R-SFFV LTR and related MuLV LTRs is presented in Fig. 8. There is greater than 95% homology between R-SFFV and the three viruses presented, F-SFFV, Friend mink cell focus-inducing leukemia virus (F-MCF), and F-MuLV. Because the tissue tropism of murine retroviruses has recently been associated (6, 22) with the U3 domain of LTRs, which also are known to contain enhancers of transcription, a search was made for homology to a putative enhancer "core sequence" (23, 24) in the U3 domain of the R-SFFV LTR. At the same relative position in the R-SFFV U3 (Fig. 8, nucleotides 147 to 153) that the Moloney murine sarcoma virus core sequences are found, there is a 5 of 7 match with the Moloney murine sarcoma virus A repeat enhancer and a 6 of 7 match with the B repeat enhancer (24). Koch et al. (22) have identified nine nucleotides in the U3 domain that are consistently different when erythrotropic MuLVs are compared with thymotropic MuLVs. In R-SFFV, an erythrotropic virus, all nine of these nucleotide differences are conserved, which supports the suggestion that they may play a role in determining tissue tropism. Whether these sequences have any functional role in R-SFFV transcription is not addressed in this report; we only note their existence.

A previous report from this laboratory showed that R-SFFV encodes a functional reverse transcriptase (4). Since

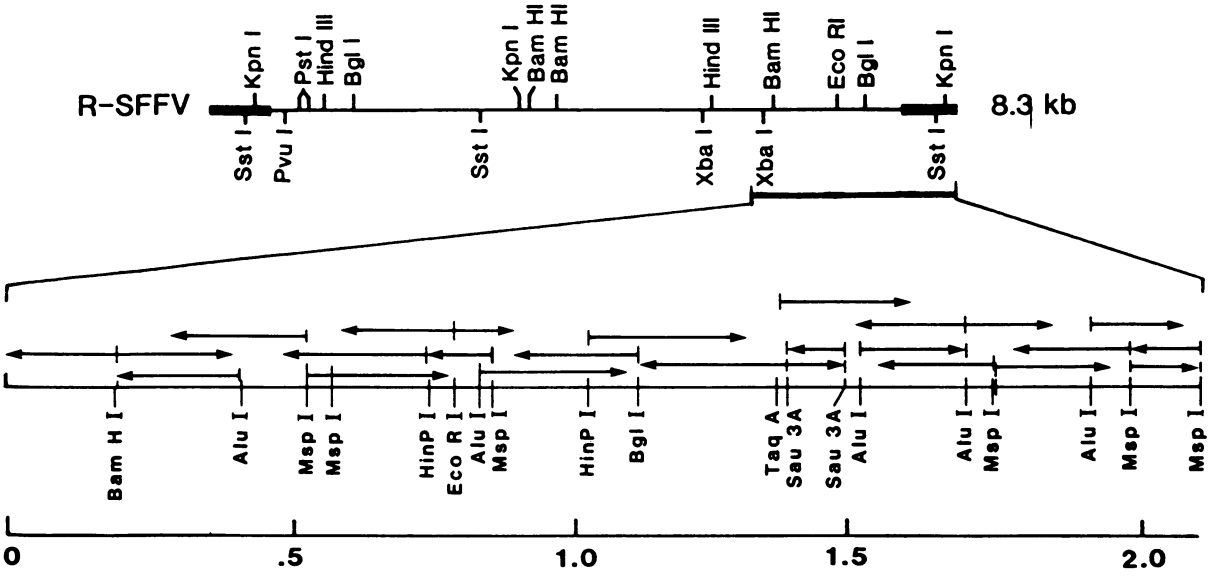

FIG. 6. Sequencing strategy for the *env*-LTR region of R-SFFV. A linear retroviral DNA structure is shown rather than the permuted structure of the plasmid clone used as a DNA source. Restriction fragments were cloned into the M13 cloning vectors MP8 and MP9 and were sequenced by the Sanger dideoxynucleotide method. The sites used are shown by vertical lines, and the arrows indicate the direction and extent of the sequencing reactions. Only the restriction sites used in generating the sequenced clones are shown, and therefore the restriction map at the bottom of the figure is not complete.

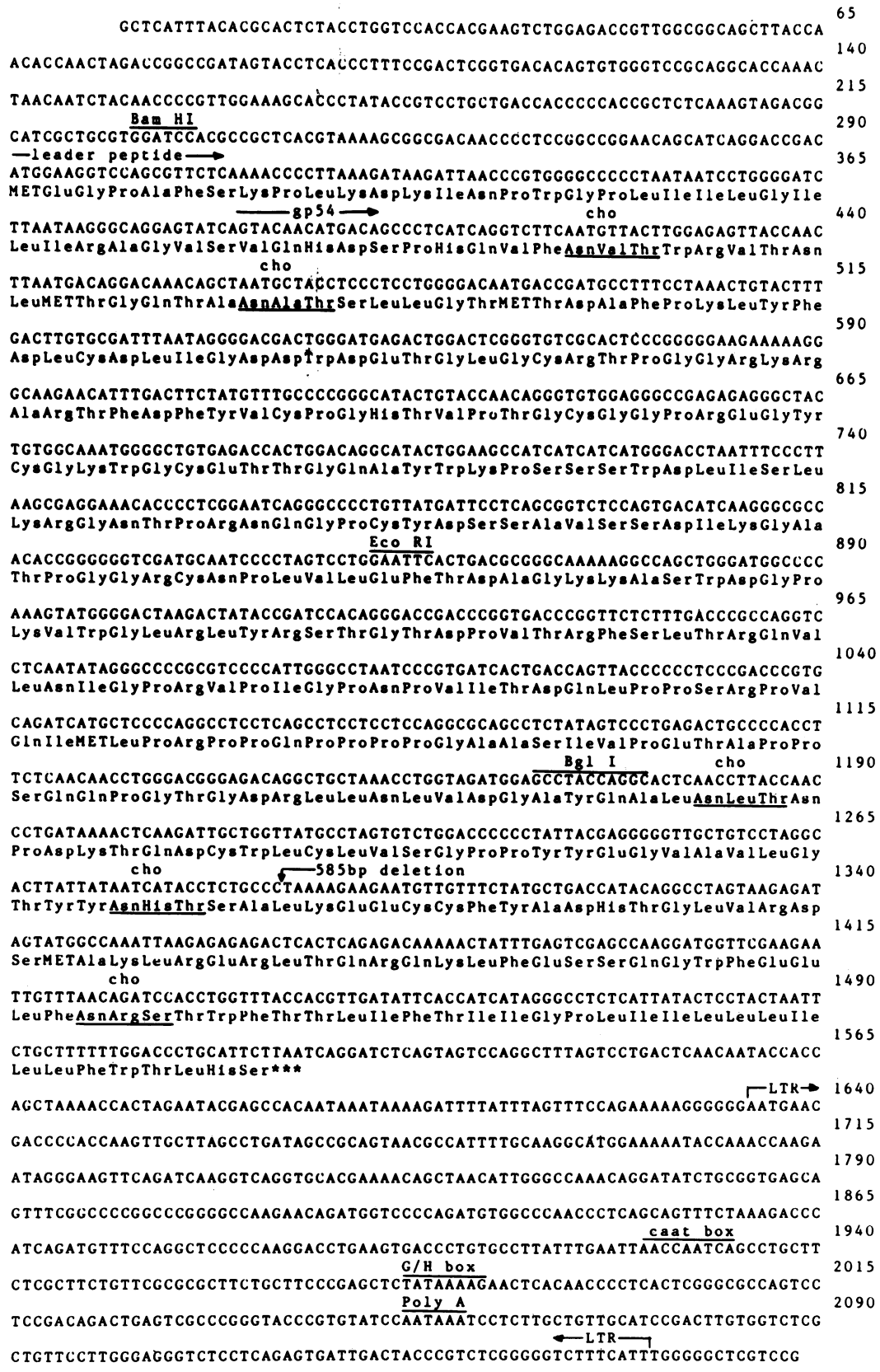

FIG. 7. Nucleotide sequence of the *env* gene and LTR of R-SFFV and the deduced amino acid sequence of gp54. Potential carbohydrate attachment sites are underlined and marked as "cho." Also marked is the deletion of 585 nucleotides of ecotropic-related sequences that result in a fusion of gp70-related sequences to p15E sequences. Within the LTR, three regions of interest have been marked: the caat box, Goldberg-Hogness (G/H) box, and the polyadenylic acid (poly A) addition signal. A 13-nucleotide imperfect terminal inverted repeat is located at the boundaries of the LTR but is not marked in this figure. The numbers on the right are nucleotide base pairs.

the sequence data covered a significant portion of the *pol* gene, a comparison of R-SFFV *pol* sequences with other MuLV *pol* genes was possible. Figure 9 shows the nucleotide sequence and the deduced amino acid sequence for the 3'-terminal 115 amino acids of the R-SFFV *pol* gene. As reported for both ecotropic (22) and xenotropic (39) MuLV *pol* genes, there is a 19-amino-acid overlap with the 5'-terminal end of the *env* gene. The R-SFFV amino acid sequence in the overlap region is identical to the xenotropic sequence, and the nucleotide sequence is a 56-of-57 match. The F-SFFV sequence reported by Wolff et al. (55) contains a 74-base-pair deletion at the 3' end of the *pol* gene, and the location of this deletion is shown in Fig. 9.

The deduced amino acid sequence for R-SFFV gp54 is shown in Fig. 7 and is compared in Fig. 10 with the three published F-SFFV polycythemic sequences (1, 8, 55), two F-MCFs (22), and F-MuLV (21). Considering the independent origin of the Rauscher and Friend viral complexes, it is remarkable that the unique structural features of the F-SFFV *env* genes are duplicated exactly in R-SFFV. These features include the location and size of the deleted ecotropic sequences that result in a fusion with p15E, the single-

```
          ┌──U3──►                                                                    (70)
1 R-SFFV  AATGAACGACCCCACCAAGTTGCTTAGCCTGATAGCCGCAGTAACGCCATTTTGCAAGGCATGGAAAAAT
2 F-SFFV  ------A-------------------------------------------------------------
3 F-MCF   ------A-------------------------------------------------------------
4 F-MULV  ------A-----------------------------T-------------------------------

                                                                                  (140)
1 ACCAAACCAAGAATAGGGAAGTTCAGATCAAGGTCAGGTGCACGAAAACAGCTAACATTGGGCCAAACAG
2 ---------------------------------G-G---A----------------G------------A
3 ---------------------------------G-G---A----------------G------------A
4 ----------------A----------------------A-----------------------------

          enhancer
1 GATATCTGCGGTGAGCAGTTTCGGCCCCGGCCCGGGGCCAAGAACAGAT
2 ------------A-------------------------------------
3 ------------A-------------------------------------
4 --------T---A-------------------------------------ACGCTGGGCCAAACAGGATAT

                                                                                  (206)
1                                                                 GGTCCCCAGATGTGGCC
2                                                                 -----------A-----
3  enhancer                                                       -----------A-----
4 CTGTGGTAAGCAGTTTCGGCCCGGTCGGCCCCGGCCCGAGGCCAAGAACGGAT-----------A-----

                                                                                  (276)
1 CAACCCTCAGCAGTTTCTAAAGACCCATCAGATGTTTCCAGGCTCCCCCAAGGACCTGAAGTGACCCTGT
2 ------------------T-----------------------------------------A---------
3 ------------------T-----------------------------------------A---------
4 ------------------T-----------------------------------------A---------

                                                                                  (346)
1 GCCTTATTTGAATTAACCAATCAGCCTGCTTCTCGCTTCTGTTCGCGCGCTTCTGCTTCCCGAGCTCTAT
2 -------------------------C-------------------------T-----------------
3 ---------A---------------C-------------------------T-----------------
4 -------------------------C-------------------------------------------

                          ◄──U3──┬──R──►                                          (415)
1 AAAAGAACTCACAACCCCTCACTCGG GCGCCAGTCCTCCGACAGACTGAGTCGCCCGGGTACCCGTGTA
2 ------G-------------------C-------------------------------------------T
3 ------G-------------------C-----------   ---------------------------
4 ------G-------------------C---------------T-----------------------A-T

                          ◄──R──┬──U5──►                                          (484)
1 TCCAATAAATCCTCTTGCTG TTGCATCCGACTTGTGGTCTCGCTGTTCCTTGGGAGGGTCTCCTCAGAG
2 C--------G----------A----------A-C----A-------A--------------------T
3 ------------------- ----------------------------------------------------
4 C--------G---------- ------------C-------------------------------------

                          ◄──U5──(516)
1 TGATTGACTACCCGTCTCGGGGGTCTTTCATT
2 ---------G---AC-----------------
3 ---------------------------------
4 ---------------------------------
```

FIG. 8. Nucleotide sequences of the LTRs of R-SFFV, F-SFFV (8), F-MCF, and F-MuLV (22) aligned to obtain maximum homology. Numbers on the right in parentheses are base pairs. Hyphens are those nucleotides which are homologous to R-SFFV. The U3, R, and U5 domains were aligned by comparisons with other LTR sequences (22). The putative enhancer elements (see text) are underlined and are aligned by analogy to their location in Moloney murine sarcoma virus (24).

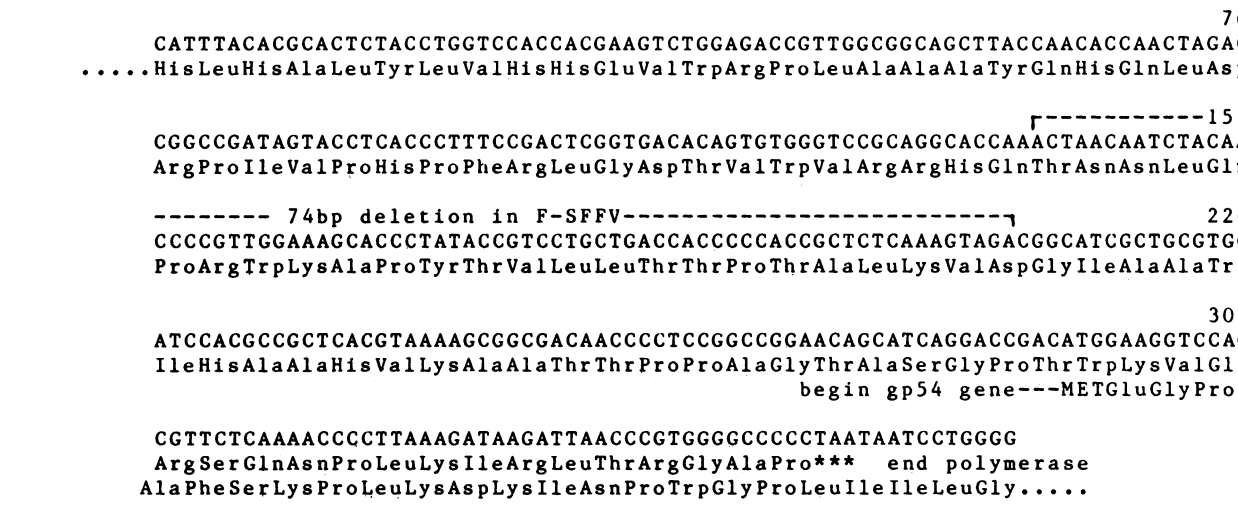

FIG. 9. Nucleotide sequence and the predicted amino acid sequences of the 3′ end of the *pol* gene and the 5′ end of the *env* gene of R-SFFV. The presence of a 19-amino-acid, frame-shifted overlap is characteristic of both xenotropic and ecotropic MuLVs (see text).

base insertion that prematurely terminates the p15E coding sequences, and the location of the 3′ xenotropic-ecotropic recombinational site. This similarity suggests that the alignment of the domains of SFFV *env* genes may be of critical importance in SFFV-induced pathogenesis. The single-base insertion of R-SFFV occurs within a stretch of six thymidine residues (nucleotides 1495 to 1500; Fig. 7), any one of which could be the added nucleotide. The additional nucleotide in F-SFFV (1, 8, 56) is within the region homologous to the stretch of six thymidines of R-SFFV. Two notable differences between F- and R-SFFV *env* genes are the absence of a six-base duplication in the hydrophobic region at the carboxy terminus of R-SFFV gp54 and the insertion of an alanine residue at amino acid 167 in R-SFFV (Fig. 10). The hydrophobic stretch of amino acids in the carboxyl-terminal, p15E-related sequences of R-SFFV gp54 (amino acids 382 to 404) has 22 of 23 hydrophobic amino acids, whereas the same region in the F-SFFV gp55s have 22 of 25 hydrophobic residues. This indicates an SFFV-specific requirement for 22 hydrophobic amino acids and could explain why F-SFFVs have acquired the six-base duplication that adds two hydrophobic leucines to this region of the glycoprotein. Perhaps the proposed membrane anchor function (26, 37) of the hydrophobic stretch has a critical requirement for a precise number of hydrophobic amino acids.

When the xenotropic-related domain of R-SFFV gp54 (amino acids 1 to 279) is compared with the same regions in the F-SFFV gp55s and to two F-MCF gp70s (22) (Fig. 10), a higher degree of amino acid homology is seen between R-SFFV gp54 and the F-MCF (98% homology) than between gp54 and F-SFFV gp55s (92% homology). This difference in homology is easily detected by visual inspection of the glycoprotein comparisons in Fig. 10. We interpret the difference in homology as evidence that R-SFFV is more closely related to the parental dual tropic MCF in the amino-terminal two-thirds of the *env* gene than are any of the F-SFFVs. It is also apparent from the differences in the three F-SFFV sequences that F-SFFVs have experienced substantial genetic drift in this region of the gene. However, several of the differences between F-SFFV and the F-MCFs (including the alanine loss at amino acid 167) are common to the F-SFFV isolates, and it is therefore likely that these were present in a common precursor of the sequenced F-SFFVs, possibly in the original F-SFFV isolate. These comparisons suggest that the xenotropic-related sequences of R-SFFV were acquired from a closely similar but possibly distinct endogenous gene from that acquired in the original F-SFFV.

The ecotropic-derived domain of gp54 starts at amino acid 280 and extends to the carboxy-terminal end of the protein. There are several consistent amino acid differences in this region between R-SFFV gp54 and all of the viruses compared with it in Fig. 10. These and other consistent differences are also apparent at the nucleotide level (comparisons not shown). We infer that the ecotropic-derived sequences in R-SFFV were probably from a different ecotropic parent (presumably R-MuLV) than those which generated the other viruses.

It is noteworthy that R-SFFV causes anemia, whereas the F-SFFVs sequenced to date cause polycythemia. Furthermore, the original F-SFFV isolate causes anemia, which suggests that the conversion to a polycythemic strain occurred after the F-SFFV origin from an F-MCF. Therefore, any amino acid differences which distinguish F-SFFVs from both R-SFFV and the MCFs may be reponsible for polycythemia. In addition, it has been reported (40a) that the difference between the polycythemia and anemia strains of F-SFFV is encoded 3′ to the *Eco*RI site at nucleotide 851 (Fig. 7). There are three amino acid differences which fit the above criteria. At both positions 337 and 357, glutamic acid residues in R-SFFV are lysines in F-SFFVs (Fig. 10). There is a corresponding net charge difference of 4 between the glycoproteins in this part of the molecule, which is in the proposed extracellular domain of p15E (26, 37). The third difference is the presence of two adjacent extra leucines in the F-SFFVs in the membrane anchor domain at the carboxyl terminus.

One of the most striking features of the SFFV-encoded *env* glycoproteins is their premature termination at an extremely hydrophobic putative membrane anchor region and the absence of any significant cytoplasmic domain. Therefore, the majority of glycoprotein molecules in infected cells must extend into the lumen of the endoplasmic reticulum and subcellular organelles, although a small proportion of the glycoprotein molecules occur on the external surfaces of the

plasma membranes (10, 40a, 41, 42). Presumably, the absence of a cytoplasmic domain capable of binding to *gag* core components may explain why gp55 is not incorporated into secreted virus particles. Other consistent structural features include a proline-rich putative hinge-type sequence (22) (amino acids 230 to 279; Fig. 10) and adjacent clusters of acidic (amino acids 83 to 87; Fig. 10) and basic (amino acids 98 to 102; Fig. 10) side chains. The striking similarities in the

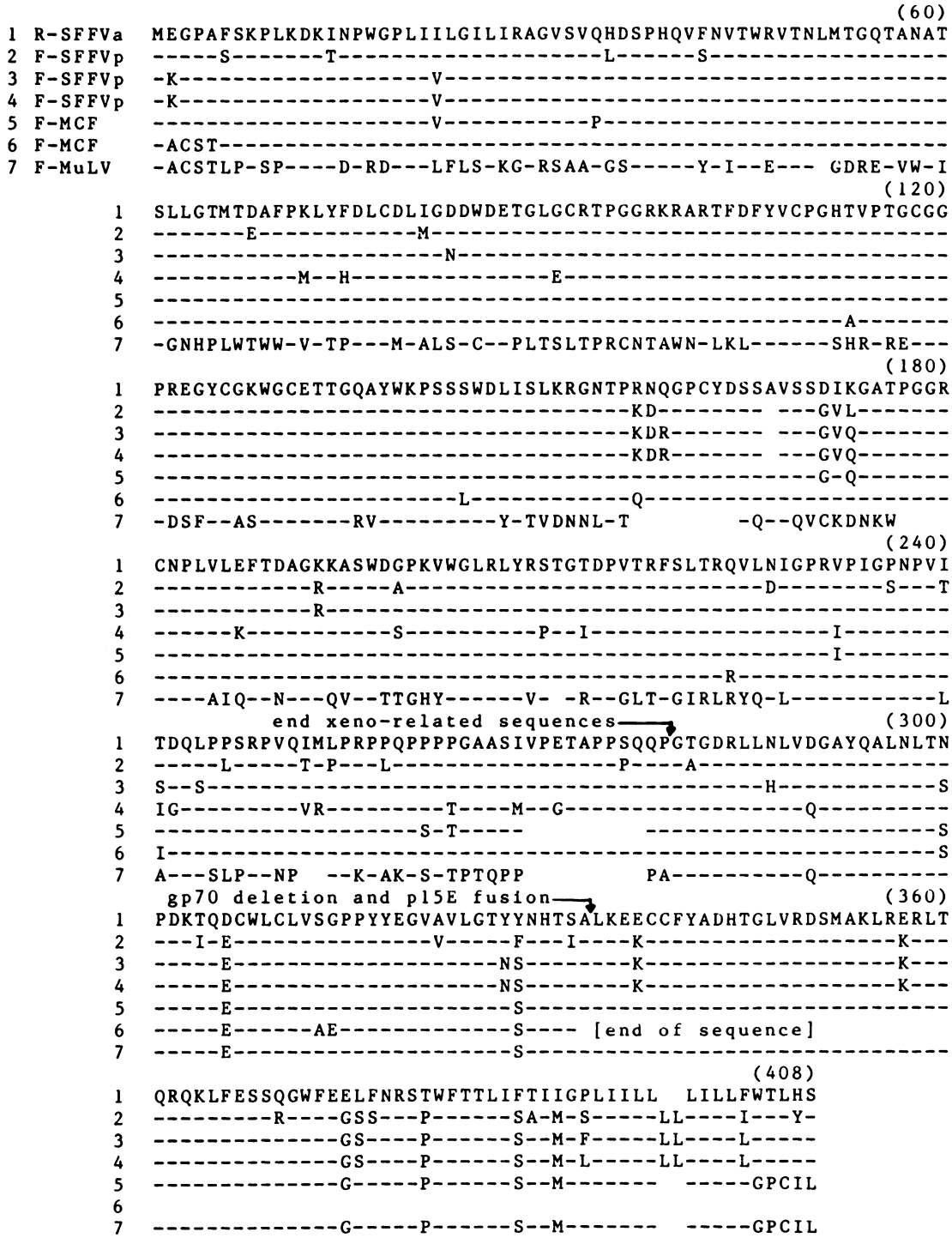

FIG. 10. Comparison of the predicted R-SFFV gp54 amino acid sequence with related MuLV *env* gene sequences. The standard one-letter abbreviations for amino acids are used. The source of these sequences are as follows: (1) R-SFFVa, this paper; (2) F-SFFVp, Wolff et al. (56); (3) F-SFFVp, Amanuma et al. (1); (4) F-SFFVp, Clark and Mak (8); (5 and 6) F-MCF, Koch et al. (21); and (7) F-MuLV, Koch et al. (22). The "a" and "p" in R-SFFVs and F-SFFVp refer to the diseases induced by these viruses, anemia and polycythemia, respectively. For F-MCF and F-MuLV, only those sequences that have homology with SFFV sequences are shown. Hyphens are those amino acids homologous to R-SFFV. The locations of the 3′ recombinational site between xenotropic- and ecotropic-related sequences and the site of fusion between gp70- and p15E-derived sequences are marked with arrows above the R-SFFV sequence.

basic domain organization of the F- and R-SFFV *env* genes implies that all regions of the encoded glycoproteins may be involved in their leukemogenic activity. Studies of mutant SFFVs are consistent with this hypothesis (30, 41).

ACKNOWLEDGMENTS

We thank Charles Faust and Curtis Machida for helpful advice and for their interest and encouragement. In addition, we thank Ethel Polonoff and Jennifer Au for expert technical assistance.

This work was supported by grant MV-160 from the American Cancer Society. R.K.B. is an N. L. Tartar Research Fellow.

LITERATURE CITED

1. **Amanuma, H., A. Katori, M. Obata, N. Sagata, and Y. Ikawa.** 1983. Complete nucleotide sequence of the gene for the specific glycoprotein (gp55) of Friend spleen focus-forming virus. Proc. Natl. Acad. Sci. U.S.A. **80**:3913–3917.
2. **Anand, R., and R. Steeves.** 1980. How many types of erythroleukemia are induced by retroviruses in mice? Nature (London) **286**:615–617.
3. **Axelrad, A. A., and R. A. Steeves.** 1964. Assay for Friend leukemia virus: rapid quantitative method based on enumeration of macroscopic spleen foci in mice. Virology **24**:513–518.
4. **Bestwick, R., M. Ruta, A. Kiessling, C. Faust, D. Linemeyer, E. Scolnick, and D. Kabat.** 1983. Genetic structure of Rauscher spleen focus-forming virus. J. Virol. **45**:1217–1222.
5. **Bosselman, R. A., J. L. D. Van Grinesven, M. Vogt, and I. M. Verma.** 1980. Genome organization of retroviruses. IX. Analysis of the genomes of Friend spleen focus-forming (F-SFFV) and helper murine leukemia viruses by heteroduplex mapping. Virology **102**:234–239.
6. **Chatis, P. A., C. A. Holland, J. W. Hartley, W. P. Rowe, and N. Hopkins.** 1983. Role for the 3′ end of the genome in determining disease specificity of Friend and Moloney murine leukemia viruses. Proc. Natl. Acad. Sci. U.S.A. **80**:4408–4411.
7. **Chattopadhyay, S. K., M. W. Cloyd, D. L. Linemeyer, M. R. Lander, E. Rands, and D. R. Lowy.** 1982. Cellular origin and role of mink cell focus-forming viruses in murine thymic lymphomas. Nature (London) **295**:25–31.
8. **Clark, S. P., and T. W. Mak.** 1983. Complete nucleotide sequence of an infectious clone of Friend spleen focus-forming provirus: gp55 is an envelope fusion glycoprotein. Proc. Natl. Acad. Sci. U.S.A. **80**:5037–5041.
9. **Cloyd, M. W., J. W. Hartley, and W. P. Rowe.** 1980. Lymphoma-genicity of recombinant mink cell focus-inducing murine leukemia viruses. J. Exp. Med. **151**:542–552.
10. **Dresler, S., M. Ruta, M. J. Murray, and D. Kabat.** 1979. Glycoprotein encoded by the Friend spleen focus-forming virus. J. Virol. **30**:564–575.
11. **Evans, L., M. Nunn, P. Duesberg, D. Troxler, and E. Scolnick.** 1980. RNAs of defective and nondefective components of Friend anemia and polycythemia virus strains identified and compared. Cold Spring Harbor Symp. Quant. Biol. **44**:823–836.
12. **Famulari, N. G.** 1983. Murine leukemia viruses with recombinant *env* genes: a discussion of their role in leukemogenesis. Curr. Top. Microbiol. Immunol. **103**:1–108.
13. **Fieldsteel, A. H., P. J. Dawson, and C. Kurahara.** 1969. Induction of lymphatic leukemia in Balb/c mice from the original isolate of Rauscher virus. Br. J. Cancer **23**:806–812.
14. **Friend, C.** 1957. Cell-free transmission in adult Swiss mice of a disease having the character of leukemia. J. Exp. Med. **105**:307–318.
15. **Fristensky, B., J. Lis, and R. Wu.** 1982. Portable microcomputer software for nucleotide sequence analysis. Nucleic Acids Res. **10**:6451–6463.
16. **Graham, F. L., and A. J. Van der Eb.** 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. Virology **52**:456–467.
17. **Haas, M., and V. Patch.** 1980. Cell-surface antigens associated with dual tropic and thymotropic murine leukemia viruses inducing thymic and nonthymic lymphomas. J. Exp. Med. **151**:1321–1333.
18. **Habara, A., E. P. Reddy, and S. A. Aaronson.** 1982. Rauscher murine leukemia virus: molecular cloning of infectious integrated proviral DNA. J. Virol. **44**:731–735.
19. **Hankins, W. D., and D. Troxler.** 1980. Polycythemia- and anemia-inducing erythroleukemia viruses exhibit differential erythroid transforming effects in vitro. Cell **22**:693–699.
20. **Hirt, B.** 1967. Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. **26**:365–369.
21. **Koch, W., G. Hunsmann, and R. Friedrich.** 1983. Nucleotide sequence of the envelope gene of Friend murine leukemia virus. J. Virol. **45**:1–9.
22. **Koch, W., W. Zimmermann, A. Oliff, and R. Friedrich.** 1984. Molecular analysis of the envelope gene and long terminal repeat of Friend mink cell focus-inducing virus: implications for the functions of these sequences. J. Virol. **49**:828–840.
23. **Laimins, L. A., P. Gruss, R. Pozzatti, and G. Khoury.** 1984. Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. **49**:183–189.
24. **Laimins, L. A., G. Khoury, C. Gorman, B. Howard, and P. Gruss.** 1982. Host-specific activation of transcription by tandem repeats from simian virus 40 and Moloney murine sarcoma virus. Proc. Natl. Acad. Sci. U.S.A. **79**:6453–6457.
25. **Larson, R., and J. Messing.** 1982. Apple II software for M13 shotgun DNA sequencing. Nucleic Acids Res. **10**:39–49.
26. **Lenz, J., R. Crowther, A. Straceski, and W. Haseltine.** 1982. Nucleotide sequence of the Akv *env* gene. J. Virol. **42**:519–529.
27. **Linemeyer, D. L., J. G. Menke, S. K. Ruscetti, L. H. Evans, and E. M. Scolnick.** 1982. Envelope gene sequences which encode the gp52 protein of spleen focus-forming virus are required for the induction of erythroid cell proliferation. J. Virol. **43**:223–233.
28. **Linemeyer, D. L., S. K. Ruscetti, J. G. Menke, and E. M. Scolnick.** 1980. Recovery of biologically active spleen focus-forming virus from moleculary cloned spleen focus-forming virus-pBR322 circular DNA by cotransfection with infectious type C retroviral DNA. J. Virol. **35**:710–721.
29. **Linemeyer, D. L., S. K. Ruscetti, E. M. Scolnick, L. H. Evans, and P. H. Duesberg.** 1981. Biological activity of the spleen focus-forming virus is encoded by a molecularly cloned subgenomic fragment of spleen focus-forming virus DNA. Proc. Natl. Acad. Sci. U.S.A. **78**:1401–1405.
30. **Machida, C. A., R. K. Bestwick, and D. Kabat.** 1984. Reduced leukemogenicity caused by mutations in the membrane glycoprotein gene of Rauscher spleen focus-forming virus. J. Virol. **49**:394–402.
31. **Mager, D., T. W. Mak, and A. Bernstein.** 1980. Friend leukemia virus-transformed cells, unlike normal stem cells, form spleen colonies in Sl/Sld mice. Nature (London) **288**:592–594.
32. **Maniatis, T., E. F. Fritsch, and J. Sambrook (ed.).** 1982. Molecular cloning. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
33. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
34. **O'Farell, P. H., E. Kutter, and M. Nakanishi.** 1980. A restriction map of the bacteriophage-T4 genome. Mol. Gen. Genet. **179**:421–435.
35. **Parker, B. A., and G. R. Stark.** 1979. Regulation of simian virus 40 transcription: sensitive analysis of the RNA species present early in infections by virus or viral DNA. J. Virol. **31**:360–369.
36. **Peschle, C., G. Migliaccio, F. Lettieri, A. R. Migliaccio, R. Ceccarelli, P. Barba, F. Titti, and G. B. Rossi.** 1980. Kinetics of erythroid precursors in mice infected with the anemia or the polycythemic strain of Friend leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **77**:2054–2058.
37. **Pinter, A., and W. J. Honnen.** 1983. Topography of murine leukemia virus envelope proteins: characterization of transmembrane components. J. Virol. **46**:1056–1060.
38. **Rauscher, F. J.** 1962. A virus-induced disease of mice characterized by erythrocytopoiesis and lymphoid leukemia. J. Natl. Cancer Inst. **29**:515–545.
39. **Repaske, R., R. R. O'Neill, A. S. Khan, and M. A. Martin.** 1983. Nucleotide sequence of the *env*-specific segment of NFS-Th-1 xenotropic murine leukemia virus. J. Virol. **46**:204–211.
40. **Ruscetti, S. K., D. Linemeyer, J. Feild, D. Troxler, and E. M.**

Scolnick. 1979. Characterization of a protein found in cells infected with the spleen focus-forming virus that shares immunological cross-reactivity with the gp70 found in mink cell focus-inducing virus particles. J. Virol. **30**:787–798.

40a. **Ruscetti, S., and L. Wolff.** 1984. Spleen focus-forming virus: relationship of an altered envelope gene to the development of a rapid erythroleukemia. Curr. Top. Microbiol. Immunol. **112**:21–44.

41. **Ruta, M., R. Bestwick, C. Machida, and D. Kabat.** 1983. Loss of leukemogenicity caused by mutations in the membrane glycoprotein structural gene of Friend spleen focus-forming virus. Proc. Natl. Acad. Sci. U.S.A. **80**:4704–4708.

42. **Ruta, M., and D. Kabat.** 1980. Plasma membrane glycoproteins encoded by cloned Rauscher and Friend spleen focus-forming viruses. J. Virol. **35**:844–853.

43. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. U.S.A. **74**:5463–5467.

44. **Scott, M. L., K. McKereghan, H. S. Kaplan, and K. E. Fry.** 1981. Molecular cloning and partial characterization of unintegrated linear DNA from gibbon ape leukemia virus. Proc. Natl. Acad. Sci. U.S.A. **78**:4213–4217.

45. **Seed, B.** 1982. Diazotizable arylamine cellulose papers for the coupling and hybridization of nucleic acids. Nucleic Acids Res. **10**:1799–1810.

46. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

47. **Southern, P. J., and P. Berg.** 1982. Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. J. Mol. Appl. Genet. **1**:327–341.

48. **Steeves, R. A.** 1975. Spleen focus-forming virus in Friend and Rauscher leukemia virus preparations. J. Natl. Cancer Inst. **54**:289–297.

49. **Teich, N., J. Wyke, T. Mak, A. Bernstein, and W. Hardy.** 1982. Pathogenesis of retrovirus-induced disease, p. 785–998. *In* R. Weiss, N. Teich, H. Varmus, and J. Coffin (ed.), RNA tumor viruses. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.

50. **Troxler, D. H., J. K. Boyars, W. P. Parks, and E. M. Scolnick.** 1977. Friend strain of spleen focus-forming virus: a recombinant between mouse type C ecotropic viral sequences and sequences related to xenotropic virus. J. Virol. **22**:361–372.

51. **Troxler, D. H., D. Lowy, R. Howk, H. Young, and E. M. Scolnick.** 1977. Friend strain of spleen focus-forming virus is a recombinant between ecotropic murine type C virus and the *env* gene region of xenotropic type C virus. Proc. Natl. Acad. Sci. U.S.A. **74**:4671–4675.

52. **Troxler, D. H., E. Yuan, D. Linemeyer, S. Ruscetti, and E. M. Scolnick.** 1978. Helper-independent mink cell focus-inducing strains of Friend murine type C virus: potential relationship to the origin of replication defective spleen focus-forming virus. J. Exp. Med. **148**:639–653.

53. **Wahl, G. M., M. Stern, and G. R. Stark.** 1979. Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl-paper and rapid hybridization by using dextran sulfate. Proc. Natl. Acad. Sci. U.S.A. **76**:3683–3687.

54. **Wendling, F., F. Moreau-Gachelin, and P. Tambourin.** 1981. Emergence of tumorigenic cells during the course of Friend virus leukemias. Proc. Natl. Acad. Sci. U.S.A. **78**:3614–3618.

55. **Wolff, L., E. Scolnick, and S. Ruscetti.** 1983. Envelope gene of the Friend spleen focus-forming virus: deletion and insertions in 3′ gp70/p15E-encoding region have resulted in unique features in the primary structure of its protein product. Proc. Natl. Acad. Sci. U.S.A. **80**:4718–4722.