## Complete nucleotide sequence of the gene for the specific glycoprotein (gp55) of Friend spleen focus-forming virus

(erythroleukemia/leukemogenic sequence/molecular structure/recombinant virus)

Hiroshi Amanuma\*, Akiko Katori\*†, Masahiro Obata‡, Noriyuki Sagata‡, and Yoji Ikawa\*‡

\*Department of Viral Oncology, Cancer Institute, Toshima-ku, Tokyo 170, Japan; and ‡Laboratory of Molecular Oncology, The Institute of Physical and Chemical Research, Wako, Saitama 351, Japan

Communicated by Susumu Ohno, March 23, 1983

ABSTRACT The complete nucleotide sequence of the gene for the specific glycoprotein (gp55) of the polycythemic strain of Friend spleen focus-forming virus (SFFV) was derived from the cloned SFFV DNA intermediate. The gp55 gene is present within 1.4 kilobases of the 5' side of the 3' long terminal repeat sequence. The open reading frame predicts the primary translation product has a total of 409 amino acids with a  $M_r$  of 44,752. Comparisons of the deduced amino acid sequence of gp55 with those of the envelope (env) gene products of murine leukemia viruses (MuLVs) revealed that gp55 is composed of three distinct regions. The amino-terminal 80% of the molecule has a high degree of sequence homology with the amino-terminal portion of the gp70 of the Moloney mink cell focus-forming virus (BALB/Mo-MCFV). This portion of the BALB/Mo-MCFV gp70 is known to be coded for by the acquired xenotropic env-like sequence. The sequence of the following 66 amino acids of gp55 is highly homologous to that of the middle portion of the p15E of Moloney MuLV (Mo-MuLV). The sequence of the carboxyl-terminal 12 amino acids is specific to gp55 and a comparison of the nucleotide sequence showed that this specific amino acid sequence is due to the presence of seven extra nucleotides compared with the sequence of the Mo-MuLV.

The spleen focus-forming virus (SFFV), which is contained in the Friend leukemia virus preparation (1), is a replication-defective murine retrovirus. When inoculated as a pseudotype, SFFV causes rapid splenomegaly and erythroleukemia in susceptible adult mice (2). Depending on the SFFV strain the erythroleukemia is accompanied by either anemia or polycythemia. SFFV is considered a member of a diverse group of the acute leukemia and sarcoma viruses of mammalian and avian origin, and it is distinguished in that it has a strict target cell specificity both *in vivo* and *in vitro* and has no unique sequence of cellular origin (3).

For the elucidation of the mechanism of erythroleukemia induction by SFFV, the identification of the leukemogenic sequence and its protein product is the first step. Two specific proteins have been shown to be coded for by SFFV. One is a gag gene-related protein, and the other is an env gene-related glycoprotein (gp55; also designated gp52) (4–7). The gag generelated protein (p45) is coded for only by a certain strain of SFFV (8). In contrast, gp55 is observed in the cells infected with any SFFV strain and has been relatively well characterized in the examination of its role in leukemogenesis (3, 9). Although the molecular size and details of the peptide map of gp55 are somewhat variable depending on the SFFV strain (10), the gp55s coded for by different SFFV strains have a common structural property: they are immunologically crossreactive with gp70, the major viral envelope glycoprotein, of the ecotropic murine leukemia virus (MuLV) (4-6). Analysis of the SFFV genome 32S RNA by molecular hybridization (11), oligonucleotide fingerprinting (12, 13), or heteroduplex formation (14) has indicated that the defective env gene region of SFFV coding for gp55 is composed of the sequence homologous to the env gene sequence of the ecotropic Friend MuLV (F-MuLV) and also of the specific sequence partially homologous to the env gene sequences of the xenotropic MuLVs. In that gp55 has both the ecotropic and xenotropic env sequences, it resembles the gp70 of the replication-competent mink cell focus-forming virus (MCFV) (15), which is the env recombinant, dualtropic MuLV and is thought to arise from the ecotropic MuLV after acquiring the xenotropic *env*-like sequence (16). It has been shown (7)that gp55 could be immunoprecipitated with the antiserum that was raised against MCFV gp70 and absorbed with the gp70 of the ecotropic MuLV. However, there are significant differences between gp55 of SFFV and gp70 of the MCFV. Although MCFV gp70 is a viral envelope glycoprotein, most gp55 are not assembled into the viral particles but remain in the membranous intracellular organelles (17).

Recently, by analyzing the biological activity of the various subgenomic DNAs made *in vitro* from the molecularly cloned SFFV DNA intermediate, Linemeyer *et al.* (18) showed that the sequence coding for gp55 is necessary for the induction of erythroleukemia by SFFV, thus providing genetic evidence for gp55 as a leukemogenic glycoprotein. We felt it essential to learn further details of the structure of gp55 to obtain a clue to dissecting the mechanism of leukemogenesis by this protein. Using the molecularly cloned SFFV DNA we determined the complete nucleotide sequence of the gene for gp55. After comparing the deduced amino acid sequence of gp55 with the sequences of the *env* gene products of various MuLVs, several unique structural features of gp55 emerged.

## MATERIALS AND METHODS

**Molecularly Cloned SFFV DNA.** Molecular cloning of the polycythemic strain of SFFV and its characterization have been described (19). Briefly, NIH 3T3 cells were infected with viruses prepared from the culture supernatant of the Friend erythroleukemia cell line (T3 K-1), which has been shown to produce an excess polycythemic strain of SFFV over helper F-MuLV (20). An unintegrated closed circular form of the SFFV DNA was isolated from the infected NIH 3T3 cells by extraction of DNA by the Hirt procedure (21). Preliminary restriction enzyme analysis revealed that *Eco*RI cleaved SFFV DNA once.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: SFFV, spleen focus-forming virus; MuLV, murine leukemia virus; F-MuLV, Friend MuLV; Mo-MuLV, Moloney MuLV; MCFV, mink cell focus-forming virus; Mo-MCFV, Moloney MCFV; bp, base pair(s); kb, kilobase(s); LTR, long terminal repeat.

<sup>&</sup>lt;sup>+</sup> On leave from the Central Research Laboratory, Hitachi, Ltd., Tokyo 185, Japan.

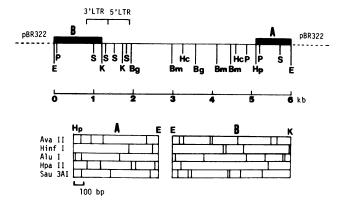


FIG. 1. Restriction enzyme map of the molecularly cloned permuted SFFV DNA (pMSF4). Bg, Bgl II; Bm, BamHI; E, EcoRI; Hc, HincII; Hp, Hpa I; K, Kpn I; P, Pvu II; and S, Sma I. bp, base pairs; kb, kilobases.

The circular SFFV DNA was thus linearized with *Eco*RI, isolated by agarose gel electrophoresis, and inserted into the plasmid vector pBR322 at the site of *Eco*RI. Several clones containing the SFFV DNA were isolated. One of them, pMSF4, was used for sequence analysis because it was shown by restriction enzyme analysis to represent the entire SFFV genome. By transfection and rescue assay (22), pMSF4 was shown to give rise to the biologically active SFFV.

DNA Sequence Analysis. Restriction enzymes were purchased from Takara Shuzo (Kyoto, Japan) or Bethesda Research Laboratories. The DNA fragments generated by digestion with Ava II, EcoRI, Hpa II, HinfI, Hpa I, Sau3AI, and Sma I were used for sequence analysis. The restriction fragments were labeled at their 5' end by using  $[\gamma^{-32}P]$ ATP (Amersham, 3,000 Ci/mmol; 1 Ci =  $3.7 \times 10^{10}$  Bq) and T4 polynucleotide kinase (Boehringer Mannheim) as described (23). End-labeled DNA fragments were digested with the appropriate restriction enzymes and separated by polyacrylamide gel electrophoresis. With some DNA fragments, strands were separated by a published method (23). The nucleotide sequence was determined by the procedure of Maxam and Gilbert (23). The sequences were confirmed by sequence analysis of the complementary strand or the fragments generated by digestion with different enzymes.

## **RESULTS AND DISCUSSION**

Nucleotide Sequence of the SFFV gp55 Gene and the Deduced Amino Acid Sequence. Fig. 1 illustrates the physical map

GICTEGAAGCLAFIGECCELTECTATACAGCLEGTEGATACAGCCEGTEGATACAGCCECTEGGETGAGCCGCEGTAGAGCCAGCCAGCAGCCACCCCCCTCCGCGCGGAGCAGCCCCCCTCGGCGGAGCCAGCC		CAGCCCCTCTCTCCAAGCT <u>CACTTACAG</u> GCCCTCCAAGCAGTACAACG	(52)
10 10 10 10 10 10 10 10 10 10	<b>4</b> Pvu II		(179)
Net lays City Pro Ate Pro Ser Lag Pro Lei Lays App Lag Lie Am Pro Trp City Pro Leu Lie Yai Leu City Lie Leu Lie Ang Ata City Val Ser Ata GAA GGT CCA GGG TH TCA AAA CCC CIT AAA GAT AAA GAT TAA CCCG TGG GGG CCC CIT A TA AFA GGG CCA GGA GAT ATA CC (402) Val Cin His App Ser Pro His Cin Yai Pre Am Val Thr Trp Ang Yai Thr Am Leu Met Thr City Cin Thr Ata Am Ata Thr Ser Leu Leu Cit Val Cin His App Ser Pro His Cin Yai Pre Am Val Thr Trp Ang Yai Thr Am Leu Met Thr City Cin Thr Ata Am Ata Thr Ser Leu Leu Cit Val Cin His App Ser Pro His Cin Yai Pre Am Val Thr Trp Ang Yai Thr Am Leu Met Thr City Cin Thr Ata Am Ata Thr Ser Leu Leu Cit Val Cin His App Ser Pro His Cin Yai Pre Am Val Thr Trp Ang Yai Thr Am Leu Met Thr City Cin Thr Ata Am Ata Thr Ser Leu Leu Cit Val Cin His App Ser Pro His Cin Yin Pre App Leu Cie Am Sta Thr Ata GGG ACA CAC CAC TA CC TA CCC CCC (498) 100 100 100 100 100 100 100 100 100 10			
40       UNU       50       CNU       50         CL CH 0       CO       CH 0       CO       CH 0       CO         CL CH 0       CL CH 0       CH	Met Lys Gly Pro Ala Phe Ser Lys Pro Leu Lys Asp Lys Ile Asn Pro Trp Gly P ATG AAA GGT CCA GCG TTC TCA AAA CCC CTT AAA GAT AAG ATT AAC CCG TGG GGC C	Pro Leu Ile Val Leu Gly Ile Leu Ile Arg Ala Gly Val CCC CTA ATA GTC CTG GGG ATC TTA ATA AGG GCA GGA GTA	A TCA
70       80       90       110       90       110       110       110       120       Small       1541         100       110       120       Small       120       Small       1541         101       120       Small       120       Small       1541         102       120       Small       120       Small       1541         103       120       Small       120       Small       1541         104       120       Small       120       Small       1561         103       120       120       Small       150       160       160         104       Small       150       160	Val Gln His Asp Ser Pro His Gln Val Phe Asn Val Thr Trp Arg Val Thr Asn L	Leu Met Thr Gly Gln Thr Ala Asn Ala Thr Ser Leu Leu	eu Gly
ACA ATG ACC GAT GCC TTT CCT AMA CTG TAC TTT GAC TTG TGC GAT TTA ATA GGG AAC GAC TGG GAT CGA ACT GGA ACT GGA CTC GGG TGT CC CACT CCC GGG 100 110 112 120 Small 120 150 150 150 150 150 150 150 150 150 15		90	,
Gig Arge Lye Arg Cit La Arge Thr Phe Arge Phe Tyr Val Cye Prio Gig Hits Thr Val Pro Tri Gig Cig Gig Cig Cig Cig Cig Cig Cig Cig Cig Cig C	ACA ATG ACC GAT GCC TTT CCT AÃA CTG TÂC TTT GAC TTG TĜC GAT TTA ATA GGĜ A	AAC GAC TGG GAT GAG ACT GGA CTC GGG TGT CGC ACT CC	C.GGG
130       140       150       150       150       160         170       Cly Cue Clu Thr Thr Gly GLA ALA TYP TTP, Use Pro Ser Ser Ser Trp Asp Leu IIe Ser Leu Lye Ang Cly Asn Thr Pro Lye GE GEC CAG GCA TCA TCA TCA TCA TCA TCA TCA TGE GAC CTA ATT TCC CTT AAG CGA GGA AAC ACT CCT AAG GAT CEG GEC CAG GCA TCA TCA TCA TCA TCA TCA TGE GAC CTA ATT TCC CTT AAG CGA GGA AAC ACT CCT AAG GAT CEG CGC (786)         170       170       170       180       190         170 Crys Tyr Asp Ser Ser Val Ser Ser Gity Val Cln Gly Ala Thr Pro Gly Gly Ang Yee Asn Pro Leu Val Leu Clu Phe Thr Asp Ala Gly Arg       190         170 Crys Tyr Asp Ser Ser Val Ser Ser Gity CGG GGT CGA TGC AC CGG GGT CGA TGC CAC CGG TCT TCA GGA TTCA CT GGA CGG GGT CA TGG CAC CT TA GGA TTC ACT GGA CGG GGT CA TGG CAC CT GGA CT CGG CT TA CGA GGA CT AGG GGT CGA TGC CGC CGG TTC TT TG GAC CGC CGG TCC CGG GT CC CGG GT CC CAA GT GGG CC CAA GT AGG GGT CAA GGA CT AGG GGT CAA GGA CT AGG GGT CAA GGA CT AGG CC CAC GG CTT CT TTG GAC CGC CGG TCC CGG GT CC CCA GG CCC GG TTC CC CAA GT AGG GGA CAA GGA TCA GC GT CAC GG GC CCA GG CCC GG TC CCC CCA GG CCC GG TC CCC CAA GGA CT AGG CCC AGG CCC GG TC CCC CAA GGA CCC GG TC CCC CCC CCC CCC CCC CCC CCC C	Gly Arg Lys Arg Ala Arg Thr Phe Asp Phe Tyr Val Cys Pro Gly His Thr Val F GGA AGA AAA AGG GCA AGA ACA TTT GAC TTC TAT GTT TGC CCC_GGG CAT ACT GTA C	Pro Thr Gly Cys Gly Gly Pro Arg Glu Gly Tyr Cys Gl CCA ACA GGG TGT GGA GGG CCG AGA GAG GGC TAC TGT GG	GÇ AAA
170       180       190         Pro Cys Tyr Asp Ser Ser Val Ser Ser Val Ser Ser Git Val Gin Giy Ala Thr Pro Giy Giy Asp Pro Leu Val Leu Clu Phe Thr Asp Ala Giy Arg       CCC TET TAT GAT TCC TCE GET CTC AGE GET GCC AGE GET GCC ACA CCE GEG GEG CCA ACC CT TEG GTC TTA GAA TTC ACT GAC GGE GEG AGA         200       210       200         Lys Ala Ser Trp Asp Giy Pro Lys Val Trp Giy Leu Arg Leu Tyr Arg Ser Thr Cly Thr Asp Pro Val Thr Arg Phe Ser Leu Thr Arg Cln Val       (882)         AGE GCC AGC TEG CAT GEG CCC AMA GTA TGE GEG CT AGA CTG GGC GGA TCC ACA GEG ACC GAC CCA GTG ACC CGE TTC TTT GAC CCE CCE GTT       (878)         Leu Asn ILe Cly Pro Arg Val Pro Lys Val Pro Asn Pro Val ILe Ser Asp Cln Ser Pro Pro Ser Arg Pro Val Clin ILe Met Leu Pro Arg Pro CTC AAT ATA GGE CCC CCC ATT GGE CCC TAT GGE CCT GTA ATC CTG GAC CAG TCA CCC CCC CCC CCC GT GTA GTA CCC GTA GTA GGE CCC CCC TCT ATA GGE CCC CCT CT CT CAA GAC CCC GC GGA GCC TCT ATA GTC CCC GAG ACC GCC CCC ATT GGE CCC CA CCC TCT ATA GTC CCC GAG ACC GGG AGC AGG CTG CTA CAC CTE GTA CCC CCC CA CCC TCT CT CAA GAC CCC CCC CC GTA TTA GGE CCC CCC TCT ATA GTC CCC CCA GCC TCT ATA GTC CCC CCA CCC TCT CT CAA GAA ACC CAA GAC CCC CCC CC CC CC TA TTA CC CCC CAA GCC TCT ATA GTC CCC CCC CAA GCC TCC CCC CAA GCC TCT ATA GTC CCC CCC CAA GCC TCT CCC CCA GCC GCA GCC CCC CCC GGE GGA GCC CCC CCC GGE GCA GCC CCC CCC GGE GCA GCC CCC CCC GC GCG GCA GCC CCC CCC	130 Two Clu Cue Clu The The Clu Cla Ala Tue Two Lus Pro See See See Tro Asp L	Leu Ile Ser Leu Lys Arg Cly Asn Thr Pro Lys Ast Ar CTA ATT TCC CTT AAG CGA GGA AAC ACT CCT AAG GAT CG	ng Gly <b>GG GGC</b>
200       210       220         Lys Ala Ser Trp Asp Gly Pro Lys Val Trp Gly Leu Arg Leu Tyr Arg Ser Thr Gly Thr Asp Pro Val Thr Ang Phe Ser CT TG ACC CGC CAG GTC       200       (978)         AG GCC AGC TGG GAT GGC CCC AMA GTA TGG GGA CTA AGA CTG TAC CGA TCC ACA GGG ACC GAC CCA GTG ACC CGG TTC TTT GACC CGC CAG GTC       (978)         Leu Ann ILe Cly Pro Arg Val Pro Ile Cly Pro Asn Pro Val ILe Ser Asp Cln Ser Pro Pro Ser Arg Pro Val Cln ILe Met Leu Pro Arg Pro CTC AAT ATA GGG CCC CGC GTC CCC ATT GGG CCT ATT CGC GGA TCC TGG GTC CCC TGT GGA CCC GTG CGA CCC GTG GTC CCC CCC CGC GTC CCC ATT GGG CCT ATT CGG GCC TAT CGG GAC CAG CCC GTG CGA GTC ATG CTC CGA GGC CCA GTC CCC CCC CCC CCC CCC CCC CCC CCC CCC	Pro Cue Tur Asp Ser Ser Val Ser Ser Clu Val Cln Clu Ala Thr Pro Clu Clu A	180 Arg Cye Aen Pro Leu Val Leu Glu Phe Thr Aep Ala Glu CGA TGC AAC CCT CTG GTC TTA GAA TTC ACT GAC GCG GG	ly Arg GT AGA
AAG GCC AGC TGE GAT GGC CCC AMA GTA TGG GGA CTA AGA CTG TAC CGA TCC ACA GGA GAC GAC CCA GGC CAC GGT TC TCT TTG ACC CGC (978)       240       250         Leu Am ILe GLY Pro Arg Val Pro ILE GLY Pro Arm Pro Val ILE Ser Arp GLN Ser Pro Pro Ser Arg Pro Val GLN ILE Met Leu Pro Arg Pro CTC ATA TA GGG CCC CGC GTC CCC ATT GGG CCT ATT CCC GTG ATC CTG GGA CCA GTC ACC CTG CGA CCC GGG CGA GAC AGG CTG CCC ATT GGC CTA ATA GGG CCC CGC GTC CCC ATT GGG CCT ATT ACT CCC GTG ATC CCT GAG ACC AGG CAG CCC CTC CCG GCG GCA GCC TCT CTA GGC GCA GCC TCT ATA GTC CCT GAG ACT GCC CCA CCT TCT CAA CAA CCT GGG AGC AGG CTG CTA CAC CTG GTA CCC CCA CCT CCT CAG CAC CTC GGG AGC AGG CTG CTA CAC CTG GTA CCC CCA CCT CCT CAA CAA CCT GGG AGC AGG CTG CTA CAC CTG GTA CAC CTG GTA ATA GTC CCT GAG ACT GCC CCA CCT TCT CAA CAA CCT GGG AGC AGG CTG CTA CAC CTG GTA CAC CTG GTA CAC GCC CCC CCT TTA GTC CCA GGC GCA GCC TCT ATA GTC CCT GAC AAA ACC CAA GAG TGC TGG TTA TGC TGG GTC GGA GAC AGG CTG CTA CAC GCC CCT ATT AGT CCT GAC CAG CCA GCC TAC CAG GCC TAC CAG GCC CTA GTA CAC GCC CAC AGT CTT GCC GAC AAA ACC CAA GAG TGC TGG TTT GCT GGA CCA CAG GAC TGA GAA ACC CAA GAG TGG TGC TG GGA CCC CCC CTA TA CAG GGA TGG TGC GGA CCA AGA GAT AGT ATT GAA ACC CAA GAA CCAA GAA TGT TGT TTT TAT GGA CCA AGA GAT AGT ATT GCA CAA AAC CTA GGC CTA GGA GCA TAG GGA TGG TTA AGA GAT AGT ATT GCA AGA CTA TTT GAA GGA GGA TGG TTG AAA AGA AAA TGT TGT	200 210	220	• •
Leu Asn ILe GLy Pro Arg Val Pro ILe GLy Pro Asn Pro Val ILe Ser Asp GLn Ser Pro Pro Ser Arg Pro Val GLn ILe Met Leu Pro Arg Pro CTC AAT ATA GGG CCC CGC GTC CCC ATT GGG CCT AAT CCC GTG ATC TCT GAC CAG TCA CCC CCC TCC CGA CCC GTG CAG ATC ATG CTC CCC AGG CT 260 270 280 (1074) Pro GLn Pro Pro Pro Pro GLy ALa ALa Ser ILe Val Pro GLu Thr ALa Pro Pro Ser GLn GLn Pro GLy Asp Arg Leu Leu His Leu Val CCT CAG CCT CCT CCT CCA GGC GCA GCC TCT ATA GTC CCT GAG ACT GCC CCA CCT TCT CAA CAA CCT GGG ACG GGA GAC AGG CTG CTA CCA CCT GG 400 (1170) 290 <b>CNO</b> 300 310 310 (1170) 290 <b>CNO</b> 300 310 310 320 GAT GGA GCC TAC CAA GCT CTC AAC AGC CTC GAC AAA ACC CAA GAG TGC TGG TTA TGC CTA GTG TCT GGA CCC CCC TAT TAC GAG GGG GTT (1266) 41a Val Leu GLy Thr Asn Ser Ala Leu Thr Ser Pro Asp Lys Thr GLn GLu Cys Trp Leu Cys Leu Val Ser GLy Pro Pro Tyr Tyr GLu GLy Val GAT GGG GTC CTA GGC ACT ATT CT AAT CAT CT TG CC CTA AAA AAA ACT CTA GTG TTG TGT TAT GC CAA GGC CTA GTA AGA GAT AGT ATG GCG GTC CTA GGC ACT AAT TCT AAT CAT ACC TCT GCC CTA AAA GAA AAAT GT TGT TTT TAT GCT GAC CAT ACA GGC CTA GTA AGA GAT AGT ATG GCC (1362) 360 370 CC MO 330 370 CC MO 380 370 CC MO 380 (1362) 198 Leu Arg Leu Thr GLn Arg GLn Lys Leu Phe GLu Ser Ser GLn GLy Trp Phe GLu GLy Ser Phe Asn Trc Ser Pro Trp Phe Thr Thr AAA TTA AGA AGA AAA CTC ATG GGA TCT CTA ATA CTT CTA TT GG GGA GGA TGG TTT AGG AGGA TCG TTT AAC AGA TCC CTG GT TTA TGC CTA GTG ATT ACC AGG TCC CTG GTT ATG CCC TGG TTA TTA CAA AGA AAA CTA TTT GGA CGA AGA CTA TTT GGA GGA TCG TTT ATG CAA GGA TCG TTT AC CAG ATG ATG GCCC CTG GTT ATT AGA AAGA A	AAG GCC AGC TGG GAT GGC CCC AAA GTA TGG GGA CTA AGA CTG TAC CGA ICC ACA G	GGG ACC GAC CCA GIG ACL CGG IIL ICI IIG ACC CGC CA	
260       270       280         Pro Gin Pro Pro Pro Pro Pro Pro Vig Ala Ala Ser Ile Val Pro Glu Thr Ala Pro Pro Ser Gin Gin Pro Gly Ang Arg Leu Leu His Leu Val       CCT CAG CCT CCT CCT CCA GGE GCA GCC TCT ATA GTC CCT GAG ACT GCC CCA CCT TCT CAA CAA CCT GGE GAG GAG AGG CTG CTA CCT GTG GTA GTG CTT CAA CAA CCC TAG GAG GAG GGE GAG ACG CTA CTA CTA CTG CTT GAG ACT GCC CCA CCT TCT CAA CAA CCC TAG GGA GCC GGA GAC AGG CTG CTA CAC CTG GAG AAA ACC CAA GAG TGC TGG TTA TGC CTA GTG TCT GGA CCC CCC TAT TAC GAG GGG GTT (170)         290       CN0       300       310       310         Asp Giy Ala Tyr Gin Ala Leu Aen Leu Thr Ser Pro Asp Lye Thr Gin Glu Cye Trp Leu Cye Leu Val Ser Gly Pro Pro Tyr Tyr Glu Cly Val GAT GGA GCC TAC CAA GCT CTC ACC AGT CCT GAC AAA ACC CAA GAG TGC TGG TTA TGC CTA GTG TCT GGA CCC CCC TAT TAC GAG GGG GTT (1266)         Ala Val Leu Gly Thr Aen Ser Aen His Thr Ser Ala Leu Lye Glu Lye Cye Cye Phe Tyr Ala Aep His Thr Gly Leu Val Arg Aep Ser Met Ala         GCG GTC CTA GGC ACT AAT TCT AAT CAT TCT GCT CTC GCC CTA AAA GAA AAA TGT TGT TTC TAT GCT GAC CAT ACA GGC CTA GTA AGA GAT AGT ATG GCC (1362)         360       370         370       CN0         380       370         390       370         390       400         390       400         409       409         Leu	Leu Asn Ile Gly Pro Arg Val Pro Ile Gly Pro Asn Pro Val Ile Ser Asp Gln S	Ser Pro Pro Ser Arg Pro Val Gin Ile Met Leu Pro Ar TCA CCC CCC TCC CGA CCC GTG CAG ATC ATG CTC CCC AG	GĞ CCT
290	Pro Gln Pro Pro Pro Pro Gly Ala Ala Ser Ile Val Pro Glu Thr Ala Pro Pro S	280 Ser Gln Gln Pro Gly Thr Gly Asp Arg Leu Leu His Le TCT CAA CAA CCT GGG ACG GGA GAC AGG CTG CTA CAC CT	eu Val TG GTA
(1266) 300 Ala Val Leu Gly Thr Aen Ser Aen His Thr Ser Ala Leu Lye Glu Lye Cye Cye Phe Tyr Ala Aep His Thr Gly Leu Val Arg Aep Ser Met Ala GCG GTC CTA GGC ACT AAT TCT AAT CAT ACC TCT GCC CTA AAA GAA AAA TGT TGT TTC TAT GCT GAC CAT ACA GGC CTA GTA AGA GAT AGT ATG GCC 360 370 Lye Leu Arg Lye Arg Leu Thr Gln Arg Gln Lye Leu Phe Glu Ser Ser Gln GLy Trp Phe Glu Gly Ser Phe Aen Arg Ser Pro Trp Phe Thr Thr AAA TTA AGA AGA GAA CAA CT ACT TT GAG TCG GAC CAA GGA TGG TTC GAA GGA TCG TTT AAC AGA TCC CCC TGG TTT ACC ACC 390 Leu Ile Ser Thr Ile Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Leu Leu Leu Trp Thr Leu His Ser END TTG ATA TCC ACC ATC ATG GGG TTT CTC ATT ATA CTC CTA CTC CTA CTA ATT CTG CTT TTA TGG ACC CTG CAT TCT TAA TCGATTAGTTCAATTHTGTAAC (1559) ACAGGGATCTCAGGTAGTCCAGGCATTAGCCCAGCTAMAACCAACCAAGAATAGGGAAGTTCAGATC CCACCAAGTTGCTTAGCCCGCAGTAACGCCATTTTGCAAGGCATGGAAMAATACCAAACCAA	Asp Gly Ala Tyr Gln Ala Leu Asn Leu Thr Ser Pro Asp Lys Thr Gln Glu Cys I	310 Trp Leu Cys Leu Val Ser Gly Pro Pro Tyr Tyr Glu Gl	<b>320</b> ly Val
Ala Val Leu Gly Thr Aen Ser Aen His Thr Ser Ala Leu Lye Glu Lye Cys Cys Phe Tyr Ala Asp His Thr Gly Leu Val Arg Asp Ser Met Ala GCG GTC CTA GGC ACT AAT TCT AAT CAT ACC TCT GCC CTA AAA GAA AAA TGT TGT TTC TAT GCT GAC CAT ACA GGC CTA GTA AGA GAT AGT AGT GGC (1362) 360 370 Lys Leu Arg Leu Thr Gln Arg Gln Lys Leu Phe Glu Ser Ser Gln Gly Trp Phe Glu Gly Ser Phe Asn Arg Ser Pro Trp Phe Thr Thr AAA TTA AGA AGA GAA CTA CTC AGA GGA CAA AAA CTA TTT GGA GGA GGC AGGA GGA TGG TTC GAA GGA TGC CCC TGG TT ACC AGA 390 400 409 Leu Ile Ser Thr ILe Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Leu Leu Leu Trp Thr Leu His Ser END TTG ATA TCC ACC ATG GGG TTT CTC ATT ATA CTC CTA CTC CTA CTA ATT CTG CTT TTA TGG ACC CTG CAT TCT TAA TCGATTAGTTCAATTGTTAAAG (1559) ACAGGATCTCAGGTGCCTGAGGCATCGACAATAACCAACC		(	
Lys Leu Arg Lys Arg Leu Thr Gin Arg Gin Lys Leu Phe Glu Ser Ser Gin Cly Trp Phe Glu Gly Ser Phe Asn Arg Ser Pro Trp Phe Thr Thr AMA TTA AGA AGA AGA CTC ACT CAG AGA CAA AMA CTA TTT GAG TCG AGC CAA GGA TGG TTC GAA GGA TCG TTT AAC AGA TCC CCC TGG TTT ACC ACG (1458) 390 Leu Ile Ser Thr Ile Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Leu Leu Trp Thr Leu His Ser END TTG ATA TCC ACC ATC ATG GGG TTT CTC ATT ATA CTC CTA CTC CTA CTA ATT CTG CTT TTA TGG ACC CTG CAT TCT TAA TCGATTAGTTCAATTGTAAAG (1559) ACAGGATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACTAACCACCAGGCATGGAAAAAAGGGCGAAGATAGGGAGAGTTCAGATC (1686)	Ala Val Leu Gly Thr Asn Ser Asn His Thr Ser Ala Leu Lys Glu Lys Cys Cys F	Phe Tyr Ala Asp His Thr Gly Leu Val Arg Asp Ser Me TTC TAT GCT GAC CAT ACA GGC CTA GTA AGA GAT AGT AT	tg gcc
Leu Ile Ser Thr Ile Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Leu Leu Leu Leu Le	Lys Leu Arg Lys Arg Leu Thr Gln Arg Gln Lys Leu Phe Glu Ser Ser Gln Gly I	Trp Phe Glu Gly Ser Phe Asn Arg Ser Pro Trp Phe Th TGG TTC GAA GGA TCG TTT AAC AGA TCC CCC TGG TTT AC	CC ACG
ACAGGATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTAAAAACCACTAGAATACGAGCCACAATAAAAAAAA	Leu Ile Ser Thr Ile Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Ile Leu L	Leu Leu Trp Thr Leu His Ser END CTT TTA TGG ACC CTG CAT TCT TAA TCGATTAGTTCAATTTGT	TTAAAG
		CCACAATAAATAAAAGATTTTATTTAGTTTCCAGAAAAAGGGGGGGAATGAA	ÁAGACĆ (1686)

FIG. 2. Nucleotide sequence of the SFFV gp55 gene and its flanking regions. Nucleotide sequence is presented as the sense (+) strand in the DNA form of the SFFV genome. The sequence is shown from the *Hpa* I site (5.1 kb in Fig. 1) to the *Sau*3AI site, which is located 290 bp upstream from the *Kpn* I site (1.2 kb in Fig. 1). The *Eco*RI site (0 and 6.0 kb in Fig. 1) is at nucleotides 862–867. The deduced amino acid sequence of gp55 is also given. Other features are: closed box, consensus 3' splice sequence; CHO, canonical sequence for glycosylation site.

of the molecularly cloned polycythemic strain of SFFV (pMSF4) based on the restriction enzyme analysis. pMSF4 had a 6.0-kb insert of the permuted SFFV DNA intermediate, which contained two long terminal repeat (LTR) sequences. It has been shown previously (13, 18) that the gp55 gene was located within 2 kb of the 5' side of the 3' LTR. The unique EcoRI site of the SFFV DNA at which the circular SFFV DNA intermediate had been linearized and inserted into the plasmid vector pBR322 was within this region, so the gp55 gene sequence was expected to separate to the two ends of the inserted SFFV DNA in pMSF4. Thus, we first isolated the Hpa I-EcoRI fragment (5.1-6.0 kb) and the EcoRI-Kpn I fragment (0-1.2 kb) of pMSF4, indicated as A and B, respectively, in Fig. 1, and determined their nucleotide sequences. Analysis of the nucleotide sequence presented in Fig. 2 showed that there is a long open reading frame (nucleotides 307-1,533) that codes for a protein having 409 amino acids. We concluded that this coding sequence is the gp55 gene. When other frames were used, the largest possible peptide coded for by those frames was <60 amino acids long.

The deduced amino acid sequence of gp55 is also presented in Fig. 2. The composition of these 409 amino acids indicates the  $M_r$  of the translation product is 44,752. About 290 bp upstream from the initiator ATG codon at nucleotides 307-309, there is a consensus 3' splice sequence (24) (7/7 match) (Fig. 2). The initiator ATG codon at nucleotides 307-309 is the first ATG codon after this splice sequence. It has been shown (8) that gp55 is synthesized from the 18S SFFV subgenomic RNA. Thus, this splice sequence is probably used to generate the 18S SFFV subgenomic RNA. Because of the extensive sequence homology with the LTRs of other murine retroviruses (25), nucleotide 1,676 can be assigned the start of the 3' LTR of SFFV (Fig. 2). This indicates that the gp55 gene is located within a 1.4-kb sequence upstream from the 3' LTR.

Because gp55 is found associated with the membranes in the SFFV-infected cells (17), it may have an amino-terminal leader

sequence that is proteolytically removed from the nascent peptide (26). Determination of the presence and the cleavage site of the leader peptide in the deduced amino acid sequence of the primary translation product of gp55 awaits the direct determination of the amino-terminal amino acid sequence of the mature gp55 by protein sequence analysis. The presence of the leader sequence means that the molecular weight of the peptide portion of the mature gp55 is slightly smaller than the value predicted for the primary translation product. We measured the molecular weight of the peptide portion of the mature gp55 by using the endoglycosidase-treated gp55 of T3 K-1 cells and it was 44,000 on polyacrylamide gel electrophoresis in the presence of NaDodSO<sub>4</sub> (unpublished data). This is in good agreement with the predicted molecular weight. The deduced amino acid sequence of gp55 shows that there are 49 basic amino acids (arginine, histidine, and lysine) and 30 acidic amino acids (aspartic acid and glutamic acid). This is consistent with the previous finding that gp55 had an isoelectric point of 8.5-9.0 (6). Near the extreme carboxyl terminus of gp55 there is a stretch of sequence (between amino acid positions 380 and 407) that is completely uncharged and mostly hydrophobic (Fig. 2). Of 28 amino acids of this sequence, 16 are either leucine or isoleucine. This hydrophobic sequence is typical of the region that anchors the membrane proteins in the lipid bilayer (26) and it probably anchors gp55 in the membranes. There are five canonical sequences, Asn-X-Thr/Ser, which can serve as glycosylation sites (27) (Fig. 2). The number of asparagine-linked oligosaccharide chains in gp55 has not yet been determined.

Comparison of the Amino Acid Sequence of gp55 with Those of the MuLV env Products. Because gp55 of SFFV has been previously shown to contain both ecotropic and xenotropic env sequences (3), it was of interest to determine how these sequences are represented in the sequence of gp55. For this purpose we compared the deduced amino acid sequence of gp55 with the amino acid sequences of the F-MuLV gp71A (28) and

SFFV gp55	M 1	K	G	P	A F	S	K	P	L 10	ĸ	DI	ĸI	N	Ρ	W	G	P 1 21		E V	L						30								40								5	50	-			-				60	Ď
Mo-MCFV <u>env</u>	M 1	Ē	G	P	A F	S	K	Ρ.	L 10	K	DI	кі	[ N	P	W	G	P 2			, L	G	I	L	IF		G 30	۷	S	vc	) H	D	S	PI	H ( 40	) )	F	N	VI		I R	V		N 1	LM	1 T	G	Q	T	A		v 1 ●60	
SFFV gp55	s	L	L	G	TM	IT	D		F 70	P	K I	LY	F	D	L	С	D 8		È G	5 N	D	W	D	E 1		i L 90	G	С	R 1	P	G	G		K F 100	R A	R	T	FI	DF	γ	٧	C 1	Р 10	GI	нт	۷	P	T	G	С	G ( 12)	
Mo-MCFV <u>env</u>	S	L	L	G	TM	IT	D	A		Ρ	KI	LY	ſF	D	L	C		Ĺ I	IG	Đ	D	W	D	EI		5 L 90	G	C	RT	P	G	G	R		R A	R	T	FI	DI	Υ	v			GI	н 1	v	Ρ	T	G	С	G ( 12)	Ğ
SFFV gp55	P	R	E	G	γc	G	ĸ		G 30	С	E	тı	G	Q	A	Y	W 14		P 5	5 S	s	W	D	LI		5 L	ĸ	R	GI	N T	P	ĸ		R (	ŞΡ	С	Y	D	s :	s -	- v	s	S 17		v	) G	A	Т	Ρ	G	G	R
Mo-MCFV <u>env</u>	P	R	E	G	YC	G	K	Ŵ		C	E	T	G	Q	A	Y		κı	P_ \$	5 5	S	W	D	LI	IS		K	R	GI	T	P	R ●	Ņ		ŠΡ	С	Y	D	S	s /	v				I   • •	G	; A	ιT	P	G	G 18	
SFFV gp55	C 180	N	Ρ	L	νı	E	F	т	D 19		G	R P	< A	s	W	D		P 1 200		/ 1	G	L	R	Ŀ١	ŕR		T 10	G	т	) F	v	т	R		S L 220	т	R	Q	V I		N I	G			v I	, I	G	; P	N	P	۷	I
Mo-MCFV <u>env</u>		N	P	L	V L	£	F	T 19	D 90		G	K ¥	( A	S	W	D		ΡI		/ 1	G	L	R	L١		2 S 210	т	G	I I ●	D F	v	т	R		Sι	τ	R	Q	V	LP	NI		23 P 30		v	5 1	( 6	i P	N	P	V 24	
F-MuLV gp71A																																		28	G T	G	D	R 29			N L	v	e Q	G	A		) # 300		, N	L	T	e N
SFFV gp55	S 240	D	Q	\$	PP	s	R	P	V 25		1	ML	. Р	R	Ρ	Ρ		P   260		<b>P</b>	G	A	A	S I	IV		E 70	т	AI	PF	, s	Q		P	ĞТ 280	G	D			Lł	ΗL	V	D 29		A				. N	L	T	S
Mo-MCFV <u>env</u>		D		L   ●	PF	S	R	P 25			I	ML	. P	R	Ρ	Ρ		PI		PF	G	A	A	S I	1 V 2		Ε	Ť	A	PF	, s	Q			GΤ	G	D	R	L	L	N L				A	1	R /	۹ L	. N	L	T 30	
Mo-MuLV <u>env</u>																																		30	ĞТ 9	Ē	N ●	R	L		N L	. V	D 32	G 20	A	1 (	ς <i>μ</i>	۹ L	. N	L	T	S
F-MuLV gp71A	P	D		T ( 31)	QE	c	W	L	C	L	۷	s e	5 P 320	P	Y	Y	E	G	v /	A 1		G 30	T	Ŷ:	s M	чн		S 37																								
SFFV gp55	Р 300	D	K	Ť	Q E	C	W	L	C 31	L	۷	s é	G P	P	Y	Y		G 1 32(		• •	Ľ	Ğ	T	N :	5 M		Ť 30	S	A	1	E	K	С	C	F Y 340	A	D	н	т	GI	LV	R	D 34	S 50	H.	A I	ĸι	LR	₹K	R	ι	т
Mo-MCFV <u>env</u>		D	ĸ	T	QE	C	W		Č 10		۲	<b>A</b> (	G P	P	Y	Y		G		Ì	L	G	т	¥ :			Ť								5-10																	
Mo-MuLV <u>env</u>	р 330	D	ĸ	T	QE	C	W		C 34		V ;	•	G P	P	Y	Y	Ē			4 1	L	G	T	¥ :			Ť	S	A 15		E	E •	C	C	FΥ	A	D		T 57		LV	R	D	S	M	AI		L R 580		R	L	N ●
SFFV gp55		R	Q	K I	LF	E	s	s			W	FE	EG	s	F	N	R	S I		a F	T	т	L	1 :	s 1			G	FI	L 1	1	L	L		ιι		L	L	L	w .	τL											
Mo-MuLV <u>env</u>	360 Q	R	Q		L F 590		s	Ţ	37 Q		W	FI		i L	F	N	R	38( S 1	0 P 1	a F	T	T 6	L 10	1 !	S 1		90 M	G	P   ●		( V 520				400							4	09									

FIG. 3. Comparison of the amino acid sequence of the gp55 of SFFV with those of the *env* products of various MuLVs. The standard one-letter amino acid code is used. The entire amino acid sequence is shown for the SFFV gp55, whereas only applicable portions are shown for the sequences of the F-MuLV gp71A (28), Mo-MuLV *env* product (29), and Mo-MCFV *env* product (31). For the SFFV gp55 and the *env* products of Mo-MuLV and Mo-MCFV, the amino acid number 1 is the amino-terminal amino acid of the primary translation product. For F-MuLV gp71A, the amino acid of the mature protein. Dots indicate positions where the amino acids are different from those of gp55.

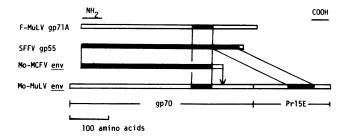


FIG. 4. Schematic representation of the results shown in Fig. 3. The closed box is the region of a high degree of sequence homology. The vertical arrow indicates the end of the substituting xenotropic *env*-like sequence in the Mo-MCFV *env* sequence. For details, see the legend to Fig. 3.

of the *env* gene products (namely, gp70 and Pr15E) of the Moloney MuLV (Mo-MuLV) (29) as ecotropic *env* sequences. F-MuLV is considered a parent virus for SFFV, and Mo-MuLV is known to be closely related to F-MuLV in the antigenic properties of its *env* products (30). We also compared the amino acid sequence of gp55 with that coded for by the substituted sequence in the BALB/Moloney MCFV (Mo-MCFV) *env* gene (31) as a xenotropic *env*-like sequence. The results of these comparisons are shown in Fig. 3 and also schematically in Fig. 4.

(i) Sequence homologous with that of the ecotropic MuLV gp70. Of the sequence of 409 amino acids of gp55, the sequence of 53 amino acids between glycine at position 279 and alanine at position 331 is highly homologous to that of the F-MuLV gp71A between positions 285 and 337. There are only four amino acid changes and no deletions (i.e., 92% homology). The sequence of these 53 amino acids of gp55 is also highly homologous (91% homology) to that of the Mo-MuLV gp70 between positions 309 and 361. In other portions of gp55 we could not find such a high degree of sequence homology with the F-MuLV gp71A or Mo-MuLV gp70. However, along the sequence of gp55 between positions 1 and 278 are scattered several stretches of sequences, each of which is significantly homologous with a portion of the F-MuLV gp71A. For example, the sequence of gp55 between positions 118 and 148 (31 amino acids) has a 77% sequence homology with that of the F-MuLV gp71A between positions 133 and 163 (data not shown).

(ii) Sequence homologous with the xenotropic env-like se-

Mo-MuLV <u>env</u>	358 Asn His Thr Ser Ala Pro Ala AAC CAT ACC T <u>CT GCT C</u> CA GCC (6848)
	559 Gly Leu Cys Ala Ala Leu Lys GGG CTG TGT G <u>CT GCT C</u> TA AAA —deletion—— (7453)
SFFV gp55	328 334 Asn His Thr Ser Ala Leu Lys AAT CAT ACC TCT GCC CTA AAA (1288) (1308)

FIG. 5. Existence of the 6-bp direct repeat sequences in the Mo-MuLV env gene in the regions that correspond to the gp70-p15E junction region of the SFFV gp55 gene. The Mo-MuLV env gene sequence is from ref. 29. The 6-bp repeat sequences are indicated by underlining.

quence. For the sequence of gp55 from positions 1 to 331, we found a very high degree of homology with the xenotropic *env*-like sequence, which is present as the substituted sequence in the Mo-MCFV gp70. Between the sequence of 331 amino acids of gp55 (positions 1–331) and the sequence of 332 amino acids of Mo-MCFV gp70 (positions 1–332), there are only 20 amino acid changes (94% homology) and 1 amino acid deletion: alanine at position 167 of the Mo-MCFV gp70 is missing in the sequence of gp55. The amino acid changes are scattered along the sequences. The sequence of Mo-MCFV gp70 between positions 1 and 332 is within the region that is coded for by the substituted sequence of the recombinant *env* gene (see Fig. 4).

(iii) Sequence homologous with that of the ecotropic MuLV p15E. For the region of gp55 between positions 332 and 409 (carboxyl-terminal amino acid) we could not find significant sequence homology with the gp70 of F-MuLV, Mo-MuLV, or Mo-MCFV. Instead, this region has a high degree of sequence homology with a portion of the p15E of Mo-MuLV. Except for the carboxyl-terminal 12 amino acids of this 78-amino acid sequence, the sequence of the preceding 66 amino acids (i.e., between positions 332 and 397) is highly homologous to the sequence of the Mo-MuLV env product between positions 557 and 622. There are seven amino acid changes (89% homology) and no deletions. The sequence between positions 557 and 622 of the Mo-MuLV env product corresponds to the sequence between residues 88 and 153 of the Pr15E, which is a cellular precursor of p15E and is 196 amino acids long (32). It was reported

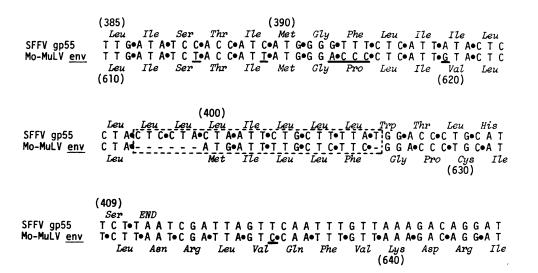


FIG. 6. Comparison of the nucleotide sequence coding for the carboxyl-terminal specific sequence of the SFFV gp55 with that of the corresponding region of the Mo-MuLV *env* gene (29). In the region indicated by the dashed line the sequence of the SFFV gp55 gene is a total of 7 bp longer than that of the Mo-MuLV *env* gene. Positions where the nucleotide differs between SFFV gp55 and Mo-MuLV *env* are indicated by underlining.

(29, 32) that the carboxyl-terminal 16 amino acids (R peptide) of the Mo-MuLV Pr15E is cleaved off during virus maturation and the remaining portion is assembled into the viral envelope as pl5E.

Only the sequence of the carboxyl-terminal 12 amino acids of gp55 (between positions 398 and 409) is specific to gp55. Other portions of gp55 are structurally closely related to the env products of the MuLVs. These sequence comparisons of gp55 with the env products of MuLVs clearly demonstrated that gp55 has a quite unique structure. First of all, it is a fusion protein composed of the gp70 portion and the p15E portion. The gp70 portion (between positions 1 and 331) is related to the xenotropic env-like sequence. Part of the gp70 portion (between positions 279 and 331) is also related to the gp70 of F-MuLV and Mo-MuLV. Compared with the general structure of the primary translation product of the MuLV env gene (i.e., NH2-gp70-Pr15E-COOH), gp55 has a deletion of the carboxyl-terminal 25% of gp70 and the amino-terminal 45% of Pr15E. The carboxyl-terminal 20% of Pr15E is also missing in gp55 and, instead, gp55 has a sequence of 12 amino acids, which is not homologous to Mo-MuLV Pr15E.

Analysis of the Nucleotide Sequence Related to the Unique Structure of gp55. The unique structural features of gp55namely, the large deletion resulting in the fusion protein and the specific sequence at the carboxyl terminus-were further analyzed at the level of nucleotide sequence. With regard to the deletion, it was of interest to learn whether there are any particular nucleotide sequences in the regions of the MuLV env gene that border the deleted sequence. As shown in Fig. 5, the sequence of the Mo-MuLV env gene (29) has a 6-bp direct repeat in these regions. If the sequence between the first nucleotide of the 5' repeat sequence and the 5' adjacent nucleotide of the 3' repeat sequence is deleted, the resulting sequence codes for the structure of the gp70-p15E junction region of gp55. These repeat sequences may have an important role in the process that results in the deletion. It would be interesting to analyze the nucleotide sequence of the env gene of F-MuLV or Friend MCFV, asking if there is a similar repeat sequence.

For the carboxyl-terminal specific sequence of gp55 we compared the nucleotide sequence of the region that codes for this sequence with that of the corresponding region of Mo-MuLV (29) (Fig. 6). In the region indicated in Fig. 6 the sequence of SFFV is a total of 7 bp longer than that of the Mo-MuLV env gene. The nucleotide sequences of SFFV preceding and following the indicated area are highly homologous to those of Mo-MuLV. Except for the sequence between positions 1,634 and 1,637, where two nucleotides are missing compared to that of Mo-MuLV (not shown in Fig. 6), the sequence of SFFV between positions 1,520 (3' adjacent nucleotide of the 3' end of the indicated area) and 1,675 (5' adjacent nucleotide of the start of 3' LTR) is co-linear and has 91% homology with that of Mo-MuLV. Despite this homology, reading frames are different between SFFV and Mo-MuLV after the indicated area in Fig. 6. With the frame of SFFV there is a termination codon (TAA) at nucleotides 1,534-1,536, whereas with the frame of Mo-MuLV a termination codon (TAG) exists 98 nucleotides downstream from that of SFFV.

How these unique structural features of gp55 are related to its cellular function in leukemogenesis is yet to be analyzed. The presence of the xenotropic env-like sequence at the aminoterminal portion resembles the structure of the MCFV gp70, which is suspected to be responsible for the induction of leukemia by this class of virus (16). A fusion protein between gp70 and p15E could result in the creation of the specific conformation in the gp55 molecule, which can make the cellular metabolism of gp55 different from that of gp70 or p15E. Finally, the absence in gp55 of the carboxyl-terminal 31 amino acids of

Pr15E may be at least partly responsible for the absence of gp55 in the viral envelope and also for its perinuclear location (17), because the presence of and the proteolytic cleavage within this sequence of the Mo-MuLV Pr15E have been suggested to be critical for the assembly of Pr15E into the viral envelope as p15E (32). To experimentally resolve the significance of these structural features of gp55 in leukemogenesis, the generation in vitro of the variant gp55 gene sequences, in which each structural feature is altered, and the test of their biological activity would be helpful.

We thank Jane Clarkin for her assistance in editing and typing the manuscript. This work was partly supported by grants from the Ministry of Education, Science and Culture and the Ministry of Health and Welfare, Japan.

- Friend, C. (1957) J. Exp. Med. 105, 307-318. 1.
- Steeves, R. A. (1975) J. Natl. Cancer Inst. 54, 289-297. 2
- Troxler, D. H., Ruscetti, S. K. & Scolnick, E. M. (1980) Biochim. 3. Biophys. Acta 605, 305-324.
- Racevskis, J. & Koch, G. (1977) J. Virol. 21, 328-337. 4.
- 5. Ikawa, Y., Yoshida, M. & Yoshikura, H. (1978) Proc. Jpn. Acad. 54, 651-656.
- Dresler, S., Ruta, M., Murray, M. J. & Kabat, D. (1979) J. Virol. 6. 30, 564-575.
- Ruscetti, S. K., Linemeyer, D., Feild, J., Troxler, D. & Scolnick, E. M. (1979) J. Virol. 30, 787–798. 7.
- Ruscetti, S., Troxler, D., Linemeyer, D. & Scolnick, E. (1980) J. 8. Virol. 33, 140-151.
- 9. Ikawa, Y., Kobayashi, Y., Obinata, M., Harada, F., Hino, S. & Yoshikura, H. (1980) Cold Spring Harbor Symp. Quant. Biol. 44, 875-885
- Yoshida, M. & Yoshikura, H. (1981) J. Gen. Virol. 52, 183-188. 10.
- Troxler, D. H., Lowy, D., Howk, R., Young, H. & Scolnick, E. 11. M. (1977) Proc. Natl. Acad. Sci. USA 74, 4671-4675.
- Evans, L. H., Duesberg, P. H., Troxler, D. H. & Scolnick, E. M. 12 (1979) J. Virol. 31, 133-146.
- Yoshida, M. & Yoshikura, H. (1980) J. Virol. 33, 587-596. 13.
- Bosselman, R. A., Van Griensven, L. J. L. D., Vogt, M. & Verma, 14. I. M. (1980) Virology 102, 234-239.
- Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. (1977) 15. Proc. Natl. Acad. Sci. USA 74, 789-792.
- Elder, J. H., Gautsch, J. W., Jensen, F. C., Lerner, R. A., Hart-ley, J. W. & Rowe, W. P. (1977) Proc. Natl. Acad. Sci. USA 74, 4676-16 4680
- Lyles, D. S. & McConnell, K. A. (1981) J. Virol. 39, 263-272. 17.
- Linemeyer, D. L., Menke, J. G., Ruscetti, S. K., Evans, L. H. 18. & Scolnick, E. M. (1982) J. Virol. 43, 223-233.
- 19. Sagata, N. & Ikawa, Y. (1982) Gan to Kagaku Ryoho 9, Suppl. 1, 290-298
- Ikawa, Y. & Yoshida, M. (1979) in Oncogenic Viruses and Host Cell 20. Genes, eds. Ikawa, Y. & Odaka, T. (Academic, New York), pp. 163-172.
- 21. Hirt, N. (1967) J. Mol. Biol. 26, 365-369.
- 22. Linemeyer, D. L., Ruscetti, S. K., Menke, J. G. & Scolnick, E. M. (1980) J. Virol. 35, 710-721
- 23. Maxam, A. M. & Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
- 24. Sharp, P. A. (1981) Cell 23, 643-646.
- 25. Van Beveren, C., Rands, E., Chattopadhyay, S. K., Lowy, D. R. & Verma, I. M. (1982) J. Virol. 41, 542-556.
- Engelman, D. M. & Steitz, T. A. (1981) Cell 23, 411-422. 26.
- 27. Neuberger, A., Gottschalk, A., Marshall, R. D. & Spiro, R. G. (1972) in Glycoproteins: Their Composition, Structure and Function, ed. Gottschalk, A. (Elsevier, Amsterdam), 2nd Ed., pp. 450-490.
- 28 Chen, R. (1982) Proc. Natl. Acad. Sci. USA 79, 5788-5792.
- 29. Shinnick, T. M., Lerner, R. A. & Sutcliffe, J. G. (1981) Nature (London) 293, 543-548.
- 30. Old, L. J., Boyse, E. A. & Stockert, E. (1964) Nature (London) **201,** 777–779.
- 31. Bosselman, R. A., Van Straaten, F., Van Beveren, C., Verma, I.
- Bossenhan, R. A., van Straaten, F., van beveren, C., verma, I.
  M. & Vogt, M. (1982) J. Virol. 44, 19–31.
  Green, N., Shinnick, T. M., Witte, O., Ponticelli, A., Sutcliffe, J. G. & Lerner, R. A. (1981) Proc. Natl. Acad. Sci. USA 78, 6023– 32. 6027.