Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 3913–3917, July 1983
Biochemistry

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* gene-related 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.

† On leave from the Central Research Laboratory, Hitachi, Ltd., Tokyo 185, Japan.

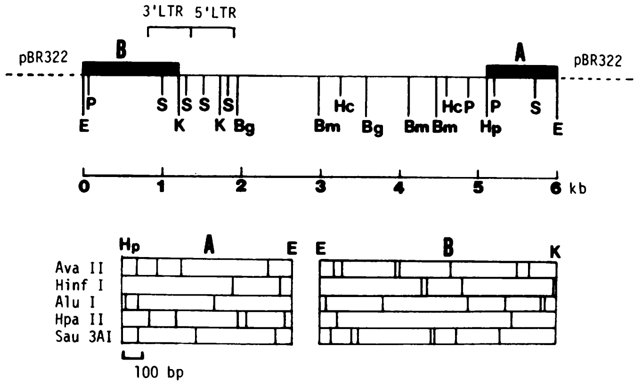

FIG. 1. Restriction enzyme map of the molecularly cloned permuted SFFV DNA (pMSF4). Bg, *Bgl* II; Bm, *Bam*HI; E, *Eco*RI; Hc, *Hinc*II; Hp, *Hpa* I; K, *Kpn* I; P, *Pvu* II; and S, *Sma* I. bp, base pairs; kb, kilobases.

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, *Eco*RI, *Hpa* II, *Hinf*I, *Hpa* I, *Sau*3AI, and *Sma* I were used for sequence analysis. The restriction fragments were labeled at their 5′ end by using [γ-^{32}P]ATP (Amersham, 3,000 Ci/mmol; 1 Ci = 3.7×10^{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

```
                                                            CAGCCCCTCTCTCCAAGCTCACTTACAGGCCCTCCAAGCAGTACAACGAGAG
                                                         (1)                                                    (52)
GTCTGGAAGCCACTGGCCGCTGCTTATCAGGACCAGCTGGATCAGCCAGTGATACCACACCCCCTCCGTGTCGGTGACGCCGTGTGGGTACGCCGGCACCAGACTAAGAACTTAGAACCCCGCTGGA
                                   ↑Pvu II                                                                             (179)
AAGGACCCTACACCGTCCTGCTGACCACCCCCACCGCTCTCAAAGTAGACGGCATCTCTGCGTGGATACACGCCGCTCACGTAAAAGCGGCGACAACCCCTCCGGCCGGAACAGCATCAGGACCGAC
                                                                                                                       (306)
1                           10                                    20                                    30
Met Lys Gly Pro Ala Phe Ser Lys Pro Leu Lys Asp Lys Ile Asn Pro Trp Gly Pro Leu Ile Val Leu Gly Ile Leu Ile Arg Ala Gly Val Ser
ATG AAA GGT CCA GCG TTC TCA AAA CCC CTT AAA GAT AAG ATT AAC CCG TGG GGC CCC CTA ATA GTC CTG GGG ATC TTA ATA AGG GCA GGA GTA TCA
                                                                                                                       (402)
                        40          CHO                     50                              CHO 60
Val Gln His Asp Ser Pro His Gln Val Phe Asn Val Thr Trp Arg Val Thr Asn Leu Met Thr Gly Gln Thr Ala Asn Ala Thr Ser Leu Leu Gly
GTA CAA CAT GAC AGC CCT CAT CAG GTC TTC AAT GTT ACT TGG AGA GTT ACC AAC TTA ATG ACA GGA CAA ACA GCT AAT GCT ACC TCC CTC CTG GGG
                                                                                                                       (498)
                    70                              80                              90
Thr Met Thr Asp Ala Phe Pro Lys Leu Tyr Phe Asp Leu Cys Asp Leu Ile Gly Asn Asp Trp Asp Glu Thr Gly Leu Gly Cys Arg Thr Pro Gly
ACA ATG ACC GAT GCC TTT CCT AAA CTG TAC TTT GAC TTG TGC GAT TTA ATA GGG AAC GAC TGG GAT GAG ACT GGA CTC GGG TGT CGC ACT CCC GGG
                                                                                                                Sma I  (594)
            100                             110                             120
Gly Arg Lys Arg Ala Arg Thr Phe Asp Phe Tyr Val Cys Pro Gly His Thr Val Pro Thr Gly Cys Gly Gly Pro Arg Glu Gly Tyr Cys Gly Lys
GGA AGA AAA AGG GCA AGA ACA TTT GAC TTC TAT GTT TGC CCC GGG CAT ACT GTA CCA ACA GGG TGT GGA GGG CCG AGA GAG GGC TAC TGT GGC AAA
                                                    ↑Sma I                                                            (690)
    130                             140                             150                                                 160
Trp Gly Cys Glu Thr Thr Gly Gln Ala Tyr Trp Lys Pro Ser Ser Ser Trp Asp Leu Ile Ser Leu Lys Arg Gly Asn Thr Pro Lys Asp Arg Gly
TGG GGC TGT GAG ACC ACT GGA CAG GCA TAC TGG AAG CCA TCA TCA TCA TGG GAC CTA ATT TCC CTT AAG CGA GGA AAC ACT CCT AAG GAT CGG GGC
                                                                                                                       (786)
                                170                             180                             190
Pro Cys Tyr Asp Ser Ser Val Ser Ser Gly Val Gln Gly Ala Thr Pro Gly Gly Arg Cys Asn Pro Leu Val Leu Glu Phe Thr Asp Ala Gly Arg
CCC TGT TAT GAT TCC TCG GTC TCC AGT GGC GTC CAG GGT GCC ACA CCG GGG GGT CGA TGC AAC CCT CTG GTC TTA GAA TTC ACT GAC GCG GGT AGA
                                                                                            Eco RI↑                    (882)
                        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 Arg Phe Ser Leu Thr Arg Gln Val
AAG GCC AGC TGG GAT GGC CCC AAA GTA TGG GGA CTA AGA CTG TAC CGA TCC ACA GGG ACC GAC CCA GTG ACC CGG TTC TCT TTG ACC CGC CAG GTC
         ↑Pvu II                                                                                                       (978)
                230                             240                             250
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 CCT
                                                                                                                      (1074)
            260                             270                             280
Pro Gln Pro Pro Pro Pro Gly Ala Ala Ser Ile Val Pro Glu Thr Ala Pro Pro Ser Gln Gln Pro Gly Thr 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 CAC CTG GTA
                                                                                                                      (1170)
    290                 CHO             300                             310                                             320
Asp Gly Ala Tyr Gln Ala Leu Asn 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 GGA GCC TAC CAA GCT CTC AAC 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)
                        CHO 330                                         340                             350
Ala Val Leu Gly Thr Asn Ser Asn His Thr Ser Ala Leu Lys Glu Lys 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 ATG GCC
                                                                                                                      (1362)
                        360                             370                         CHO     380
Lys Leu Arg Lys 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 AAG AGA CTC ACT CAG AGA CAA AAA CTA TTT GAG TCG AGC CAA GGA TGG TTC GAA GGA TCG TTT AAC AGA TCC CCC TGG TTT ACC ACG
                                                                                                                      (1458)
                390                             400                             409
Leu Ile Ser Thr Ile Met Gly Phe Leu Ile Ile Leu Leu Leu Leu Leu Ile 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 TCGATTAGTTCAATTTGTTAAAG
                                                                                                                      (1559)
ACAGGATCTCAGTAGTCCAGGCTTTAGTCCTGACTCAACAATACCACCAGCTAAAACCACTAGAATACGAGCCACAATAAATAAAAGATTTTATTTAGTTTCCAGAAAAAGGGGGGAATGAAAGACC
                                                                                                                |     (1686)
CCACCAAGTTGCTTAGCCTGATAGCCGCAGTAACGCCATTTTGCAAGGCATGGAAAAATACCAAACCAAGAATAGGGAAGTTCAGATC                        |
                                                                                      (1774)                    └ 3' LTR
```

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 *Eco*RI 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–*Eco*RI fragment (5.1–6.0 kb) and the *Eco*RI–*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 $NaDodSO_4$ (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

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 number 1 is the amino-terminal amino acid of the mature protein. Dots indicate positions where the amino acids are different from those of gp55.

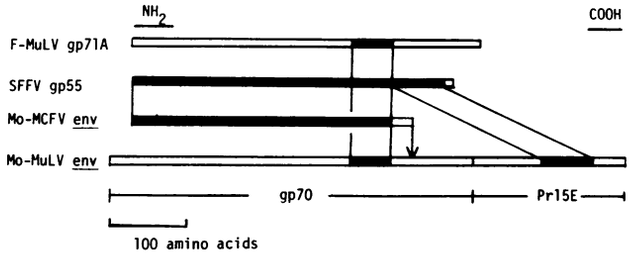

FIG. 4. Schematic representation of the results shown in Fig. 3. The closed box is the region of a high degree of sequence homology. The vertical arrow indicates the end of the substituting xenotropic *env*-like sequence in the Mo-MCFV *env* sequence. For details, see the legend to Fig. 3.

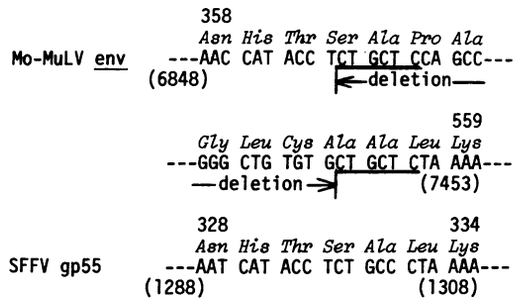

FIG. 5. Existence of the 6-bp direct repeat sequences in the Mo-MuLV *env* gene in the regions that correspond to the gp70–p15E junction region of the SFFV gp55 gene. The Mo-MuLV *env* gene sequence is from ref. 29. The 6-bp repeat sequences are indicated by underlining.

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 sequence.* 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

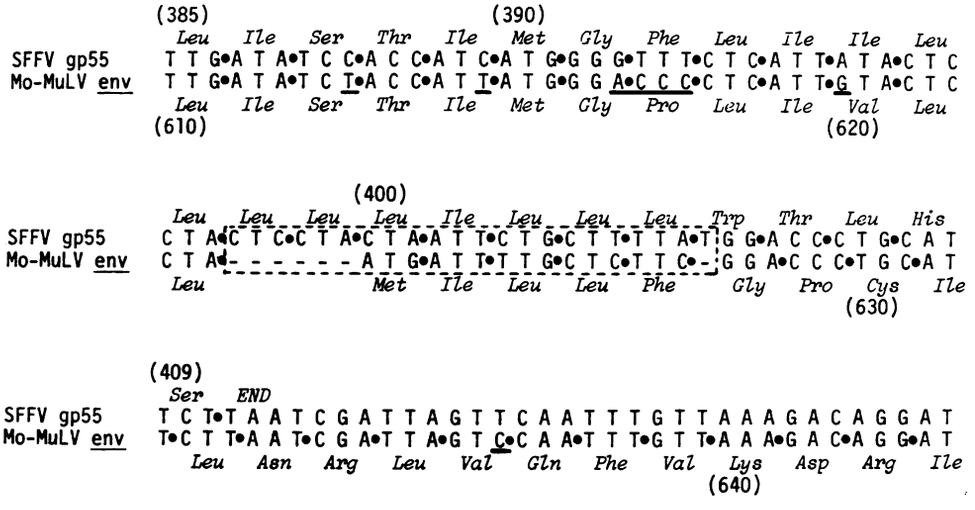

FIG. 6. Comparison of the nucleotide sequence coding for the carboxyl-terminal specific sequence of the SFFV gp55 with that of the corresponding region of the Mo-MuLV *env* gene (29). In the region indicated by the dashed line the sequence of the SFFV gp55 gene is a total of 7 bp longer than that of 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 p15E.

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., NH_2-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 gp55—namely, 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 amino-terminal 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.

1. Friend, C. (1957) *J. Exp. Med.* **105,** 307–318.
2. Steeves, R. A. (1975) *J. Natl. Cancer Inst.* **54,** 289–297.
3. Troxler, D. H., Ruscetti, S. K. & Scolnick, E. M. (1980) *Biochim. Biophys. Acta* **605,** 305–324.
4. Racevskis, J. & Koch, G. (1977) *J. Virol.* **21,** 328–337.
5. Ikawa, Y., Yoshida, M. & Yoshikura, H. (1978) *Proc. Jpn. Acad.* **54,** 651–656.
6. Dresler, S., Ruta, M., Murray, M. J. & Kabat, D. (1979) *J. Virol.* **30,** 564–575.
7. Ruscetti, S. K., Linemeyer, D., Feild, J., Troxler, D. & Scolnick, E. M. (1979) *J. Virol.* **30,** 787–798.
8. Ruscetti, S., Troxler, D., Linemeyer, D. & Scolnick, E. (1980) *J. 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.
10. Yoshida, M. & Yoshikura, H. (1981) *J. Gen. Virol.* **52,** 183–188.
11. Troxler, D. H., Lowy, D., Howk, R., Young, H. & Scolnick, E. M. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4671–4675.
12. Evans, L. H., Duesberg, P. H., Troxler, D. H. & Scolnick, E. M. (1979) *J. Virol.* **31,** 133–146.
13. Yoshida, M. & Yoshikura, H. (1980) *J. Virol.* **33,** 587–596.
14. Bosselman, R. A., Van Griensven, L. J. L. D., Vogt, M. & Verma, I. M. (1980) *Virology* **102,** 234–239.
15. Hartley, J. W., Wolford, N. K., Old, L. J. & Rowe, W. P. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 789–792.
16. Elder, J. H., Gautsch, J. W., Jensen, F. C., Lerner, R. A., Hartley, J. W. & Rowe, W. P. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 4676–4680.
17. Lyles, D. S. & McConnell, K. A. (1981) *J. Virol.* **39,** 263–272.
18. Linemeyer, D. L., Menke, J. G., Ruscetti, S. K., Evans, L. H. & Scolnick, E. M. (1982) *J. Virol.* **43,** 223–233.
19. Sagata, N. & Ikawa, Y. (1982) *Gan to Kagaku Ryoho* **9,** Suppl. 1, 290–298.
20. Ikawa, Y. & Yoshida, M. (1979) in *Oncogenic Viruses and Host Cell 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.
26. Engelman, D. M. & Steitz, T. A. (1981) *Cell* **23,** 411–422.
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. M. & Vogt, M. (1982) *J. Virol.* **44,** 19–31.
32. Green, N., Shinnick, T. M., Witte, O., Ponticelli, A., Sutcliffe, J. G. & Lerner, R. A. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6023–6027.