JOURNAL OF VIROLOGY, May 1989, p. 2379–2384
0022-538X/89/052379-06$02.00/0
Copyright © 1989, American Society for Microbiology
Vol. 63, No. 5

Nucleotide Sequence and Distinctive Characteristics of the *env* Gene of Endogenous Feline Leukemia Provirus

D. VINOD KUMAR,[1] BRIAN T. BERRY,[1] AND PRADIP ROY-BURMAN[1,2]*

Departments of Pathology[1] and Biochemistry,[2] University of Southern California School of Medicine, Los Angeles, California 90033

Received 21 November 1988/Accepted 18 January 1989

Nucleotide sequence analysis of the *env* gene of two different endogenous feline leukemia virus (FeLV) loci, CFE-6 and CFE-16, of domestic cats revealed the following characteristics. (i) Both proviruses contain an open reading frame in the *env* region; (ii) whereas the full complement of the exogenous FeLV *env* is generally present in CFE-6 DNA, it is truncated in CFE-16 DNA such that the 5′ half of the gp70 domain and the untranslated region 3′ to the p15E domain have been fused by an internal deletion, resulting in loss of the C-terminal half of the gp70- and all of the p15E-coding sequences; (iii) endogenous *env* is highly homologous to large sequence domains conserved in all three exogenous FeLV subgroups (A, B, and C) but is similar to FeLV-B sequence domains in the variable regions detected in these viruses; and (iv) there are four other sequence domains, one residing at the C terminus of gp70 and three scattered in p15E, which are unique for the endogenous *env*, thereby distinguishing it from the FeLV-B gene.

Feline leukemia viruses (FeLVs) are chronic leukemogenic retroviruses that induce neoplasms with long periods of latency. Most of the events that take place and interplay during this latency period are currently undefined, except that in feline T-cell lymphosarcomas the c-*myc* proto-oncogene is either frequently insertionally activated by an FeLV provirus or incorporated into the viral genome and transduced (5, 10, 12, 13). However, it is unknown what FeLV subgroups or variants might have the viral determinants to induce cell-specific lymphoproliferative disorders. This issue is important, especially since evidence has been obtained to suggest that specific FeLV variants may induce aplastic anemia or fetal immunodeficiency diseases in experimental cats (3, 7, 9, 16, 18). In analogy to recombinational events that give rise to tumorigenic polytropic murine leukemia viruses (6, 11, 26), it is possible that FeLV infection also results in recombination with transcriptionally active endogenous FeLV elements present in the cat genome (1, 23, 24). The hybrid virus may have altered target cell specificity and may be the precursor for pathogenesis. In fact, the possibility raised earlier that FeLV subgroup B isolates might have arisen by genetic exchanges between FeLV subgroup A and endogenous proviral elements of the normal cat genome (4, 19, 25) has found new experimental support. These recent experiments suggest transduction of endogenous FeLV *env* gene regions from feline cells by infectious FeLV subgroup A (17). The recombinant viruses thus isolated resemble FeLV subgroup B in their extended host range properties. The characterization of these chimeric FeLVs was, however, based primarily on hybridization to an endogenous FeLV *env*-specific DNA probe (25), and the detailed genetic structure of the *env* gene of the various endogenous FeLV loci of the domestic cat has not yet been reported. In this paper, we present for the first time the nucleotide sequence of the *env* gene of two independent endogenous FeLV loci. The sequence data revealed the open reading frame of this endogenous viral gene and pointed to regions of significant divergence from the exogenous viral counterpart.

Two endogenous FeLV clones, CFE-6 and CFE-16 (24), were selected for analysis. The CFE-6 clone represented the major class of nearly full-length proviruses, whereas CFE-16 belonged to the minor class of truncated proviruses present in the uninfected cat genome. The nucleotide sequence was determined by the M13 dideoxynucleotide-chain termination method (2, 20). For clone CFE-6, the 2.85-kilobase *Pst*I-*Eco*RI fragment encompassing the *env* gene and the 3′ long terminal repeat (LTR) was cleaved internally by *Hin*dIII, and each of the two ensuing fragments (1.25 and 1.6 kilobases) was cloned into both M13mp18 and M13mp19 vectors. For CFE-16 DNA sequence analysis, the 1.33-kilobase *Pst*I-*Kpn*I truncated *env* region was cloned into the same vectors. Figure 1 shows the nucleotide and deduced amino acid sequences for both cloned endogenous genes. Assignment of the functional domains for the signal peptide and for the gp70 and p15E proteins was based on similarities to reported feline leukemia virus sequences (4, 14, 19, 25).

Several interesting observations emerged from analysis of the sequence data. First, the *env* region of the CFE-6 DNA exhibited a fully open reading frame comprising 2,007 base pairs. This was in contrast to the detection of frameshift and nonsense mutations in the *gag* region of this nearly full-length endogenous provirus sequence (1). The endogenous provirus CFE-16 also contained an open reading frame of 819 base pairs in the *env* gene; however, a large deletion, extending from around the middle region of gp70 to the purine-rich region preceding the 3′ LTR, indicated that this provirus was truncated not only in the *pol* gene, as described earlier (24), but also in the *env* gene. Comparison of the nucleotide sequence of *env* retained in CFE-16 with the sequence of CFE-6 showed a very high degree of sequence conservation (approximately 99%), although a few nucleotide substitutions did occur. Most of these changes were in either position 1 or position 2 of the codon, resulting in amino acid substitutions. Altogether, there were nine amino acid changes: one was located in the signal peptide, six were scattered in the N-terminal half of gp70, and two, occurring near the C terminus of p15E, resulted from the internal deletion and consequent frameshift mutations. Sequence comparison of the LTR region immediately 3′ to the *env*

* Corresponding author.

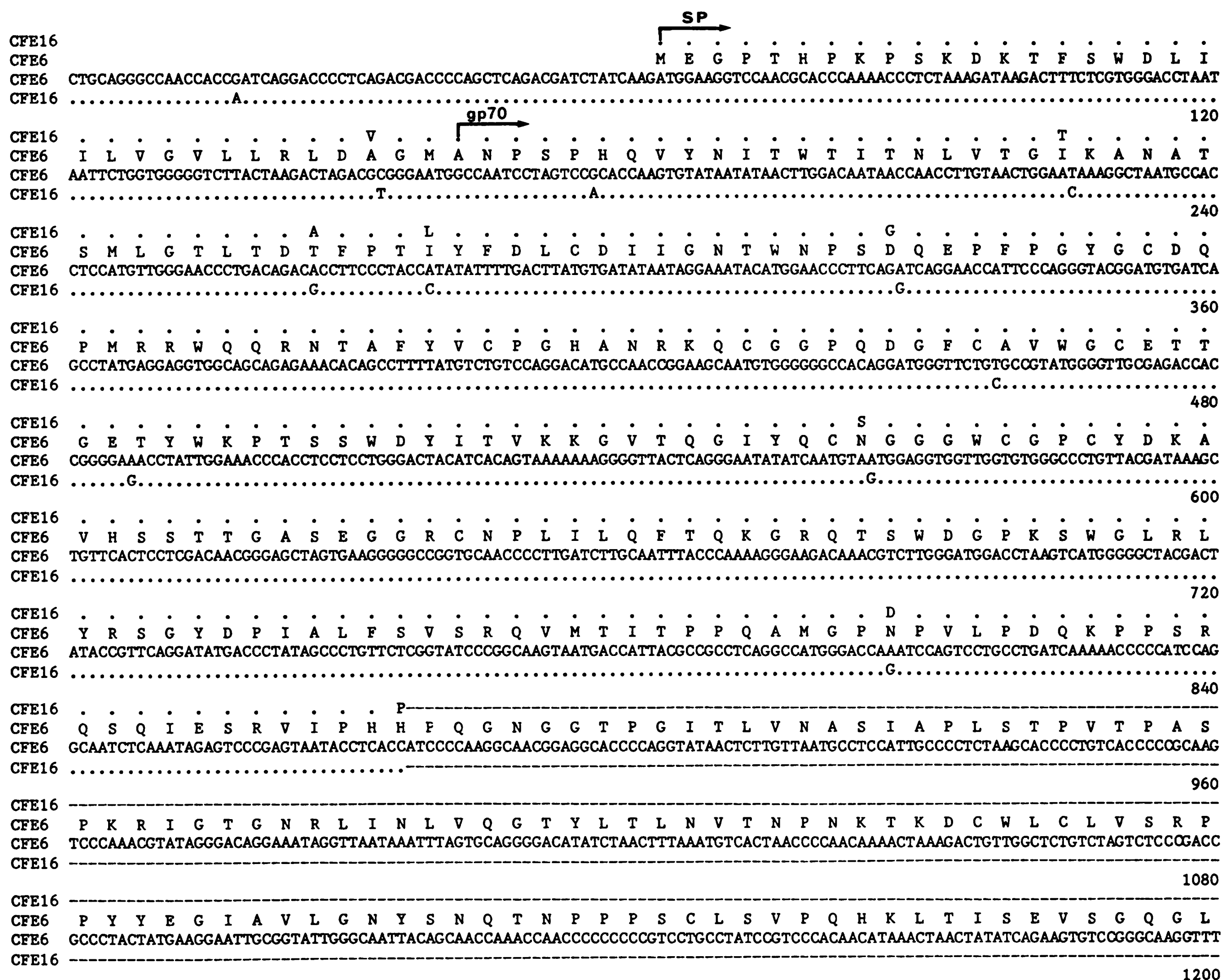

FIG. 1. Nucleotide and deduced amino acid sequences of the *env* gene of the endogenous provirus clones CFE-6 and CFE-16. The nucleotide sequence begins at the *Pst*I site 61 nucleotides upstream of the coding region and extends into the 3′ LTR. The sequence of the CFE-6 clone is shown, and the identical nucleotide or amino acid residues in the sequence of CFE-16 are indicated by dots. Changes in nucleotides and deduced amino acids in CFE-16 are shown by letters on the bottom and top lines, respectively. Deletions are marked by dashes, and the translation termination codon is indicated by the asterisk. Other features illustrated are beginning points for the signal peptide (SP), gp70, p15E, and the 3′ LTR. Also highlighted within the 3′ LTR are the inverted repeats (IR) delineating the LTR, the beginning of U3, R, and U5 sequences, and the regulatory domains such as CCAAT, TATAA, and poly(A) signal boxes.

gene showed that these LTRs were highly homologous since there were only 13 base changes, including deletions within this 697-base-pair region of CFE-6 and CFE-16. The 3′ LTR sequence of these endogenous proviruses matched well with the 5′ LTR sequences reported earlier (1), indicating potential conservation of transcription-regulatory elements in both LTRs flanking the endogenous proviral genes.

To determine the extent of amino acid sequence homology between endogenous and exogenous FeLV *env* genes, we compared the endogenous sequences with exogenous sequences representing three subgroups (A, B, and C) of FeLV classified on the basis of neutralization and viral interference (21, 22). Among the four isolates included in the comparison, ST-FeLV-B and GA-FeLV-B were from subgroup B (4, 14), FeLV-A/Glasgow was from subgroup A (25), and FeLV-C/Sarma was from subgroup C (19) (Fig. 2). This comparison clearly revealed that the endogenous *env* was more homologous to the FeLV-B gene than to either the FeLV-A or FeLV-C gene. Throughout the 662 amino acids of the gp70 and p15E regions in GA-FeLV-B, there were only 70 changes, including 69 amino acid substitutions and 1 deletion in clone CFE-6. Thus, a high degree of homology (89.5%) exists between the endogenous provirus and the exogenous FeLV-B virus at the amino acid level. Although all FeLV *env*-encoded proteins contained sequence blocks with nearly complete homology, in several regions FeLV-A and FeLV-C differed from the group represented by FeLV-B and the endogenous virus. The overall homology, considering the amino acid substitutions, deletions, and insertions, was only 73% between CFE-6 and the subgroup A or C *env* gene. The observed higher level of sequence identity between the endogenous and FeLV-B *env* genes was consistent with the

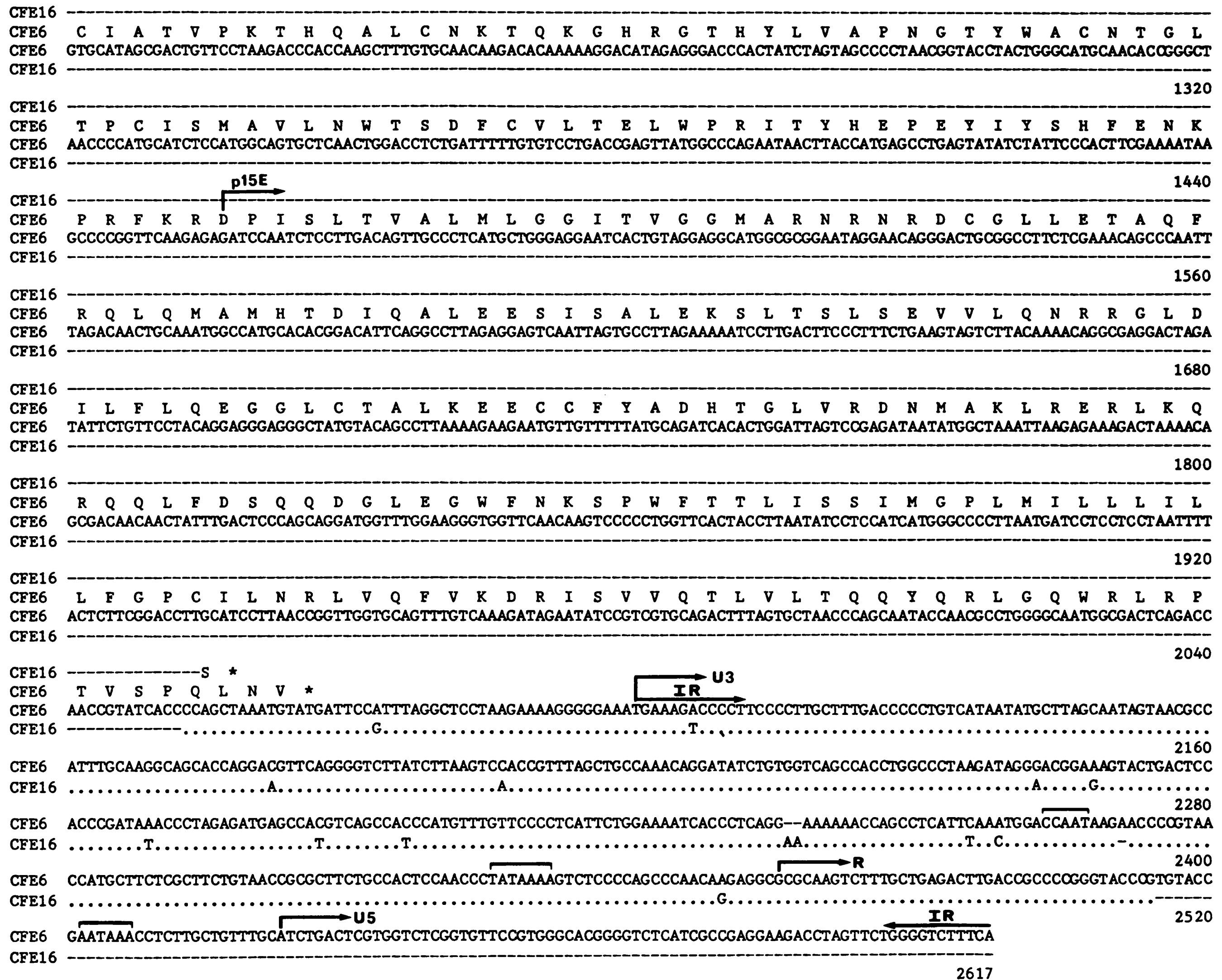

FIG. 1—*Continued.*

earlier proposal that FeLV-B isolates were derived through recombination between FeLV-A and endogenous *env* genes (19, 25). Moreover, the data supported the recent findings of transduction of endogenous sequences from feline cells by transfected FeLV-A DNA to yield FeLV-B-type viruses (17). It is difficult to assess from our sequence data where in the viral genome the hypothetical recombinational event(s) which gave rise to FeLV-B took place. It would be necessary to extend the sequencing to the adjacent *pol* gene to determine whether CFE-6 or CFE-16 diverges from FeLV-B at a discrete site in that region.

The occurrence of several variable regions of homology in the FeLV gp70 glycoprotein sequence was described earlier by different investigators (14, 19). Altogether, seven major regions of variability (I to VII in Fig. 2) were identified. It could be clearly seen that the deduced amino acid sequence of the endogenous gp70 was very similar to that of FeLV-B and not to that of FeLV-A or FeLV-C in regions II, III, IV, and V. In the 33-amino-acid length of the signal peptide (region I), GA-FeLV-B varied from ST-FeLV-B in nine amino acids, whereas there was only a single amino acid difference between GA-FeLV-B and FeLV-A in this region. The variation in signal peptide sequence between the two subgroup B isolates of the exogenous virus was, however, contrasted with the detection of homology in 32 of 33 amino acids between the corresponding sequences of CFE-16 and CFE-6 endogenous clones. Furthermore, the endogenous signal peptide region was closer in sequence to that of ST-FeLV-B (94 to 97% homology) than to that of GA-FeLV-B (73% homology), which implies diversity in recombinational processes that might have resulted in the genesis of subgroup B viruses.

Although the endogenous and FeLV-B gp70 sequences were very similar and all positions of the glycosylation sites were conserved between these sequences, the C-terminus sequence of the endogenous gp70 appeared to diverge considerably from that of the FeLV-B gp70. This variation could be seen in region VII (Fig. 2), which encompassed nine amino acids, including the first one of p15E. All exogenous FeLVs retained homology in five to eight amino acids, whereas CFE-6 differed at seven of nine positions. Accordingly, this region was fairly characteristic for the endogenous *env* gene.

Three other predicted protein domains (VIII, IX, and X in

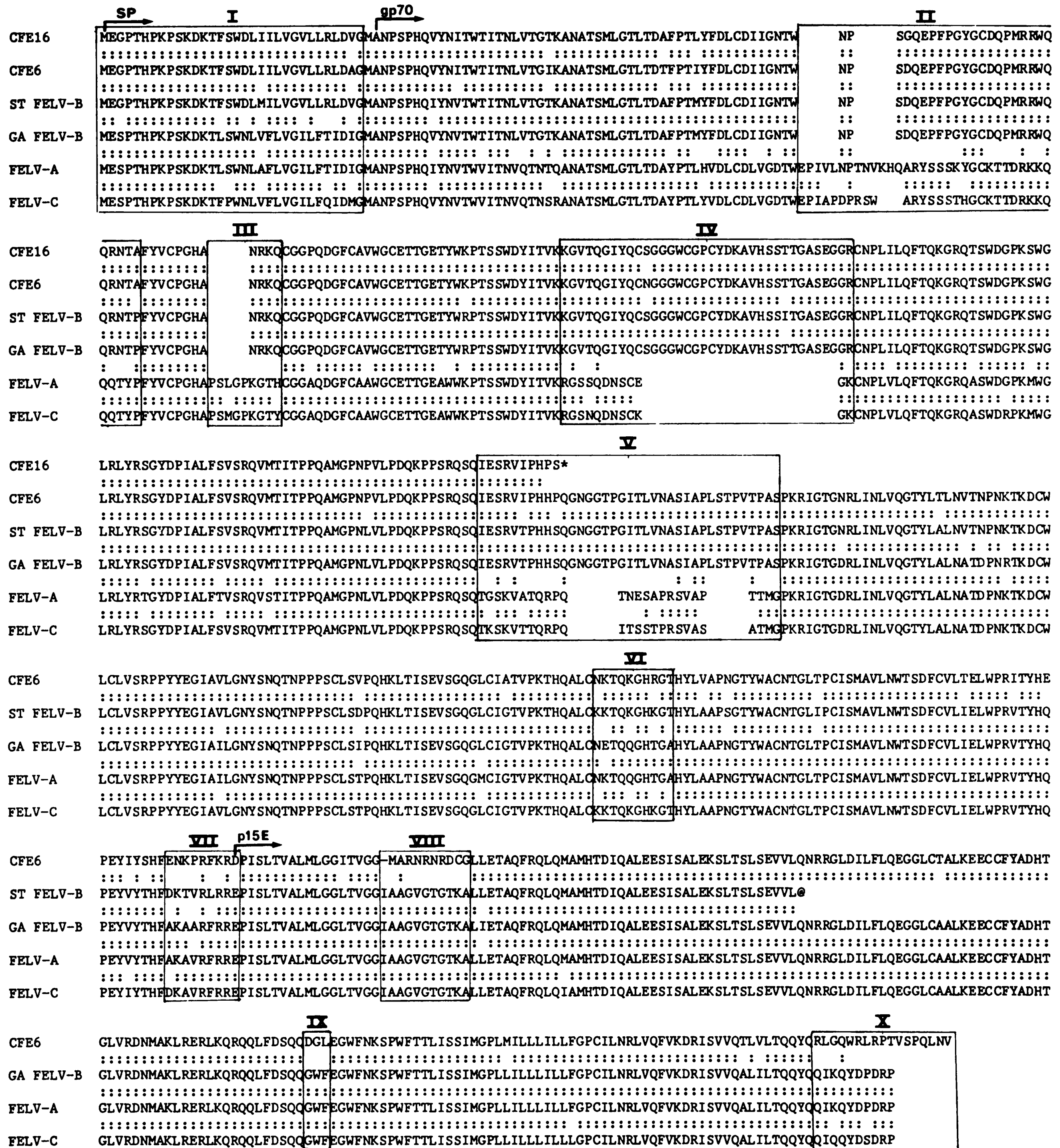

FIG. 2. Deduced amino acid sequence comparison of the *env* genes of the two endogenous proviruses with those of four exogenous FeLV isolates, ST-FeLV-B (14), GA-FeLV-B (14), FeLV-A/Glasgow (25), and FeLV-C/Sarma (19). Identical amino acids (:) and the abrupt termination in CFE-16 *env* resulting from internal deletion and consequent frameshifting (*) are indicated. The reported ST-FeLV-B sequence was that of a defective viral genome with premature termination (@) as well as deletion of the C-terminal portion of p15E (14). Boxed areas, designated I through X, indicate the major regions of variable homology. Features marked by arrows are the beginnings of signal peptide (SP), gp70, and p15E sequences. Gaps in the sequences were introduced to maximize homologies between isolates.

Fig. 2), occurring in the p15E gene region, appeared unique for the endogenous *env*. The stretch of 11 amino acids in region VIII, strongly conserved in exogenous FeLVs, was entirely different in CFE-6 except for a single amino acid. This region also contained an amino acid deletion in CFE-6. Similarly, region IX, consisting of three amino acids, was unique to CFE-6. Region X of exogenous p15E contained 10 amino acids, identical to those in FeLV-B and FeLV-A but

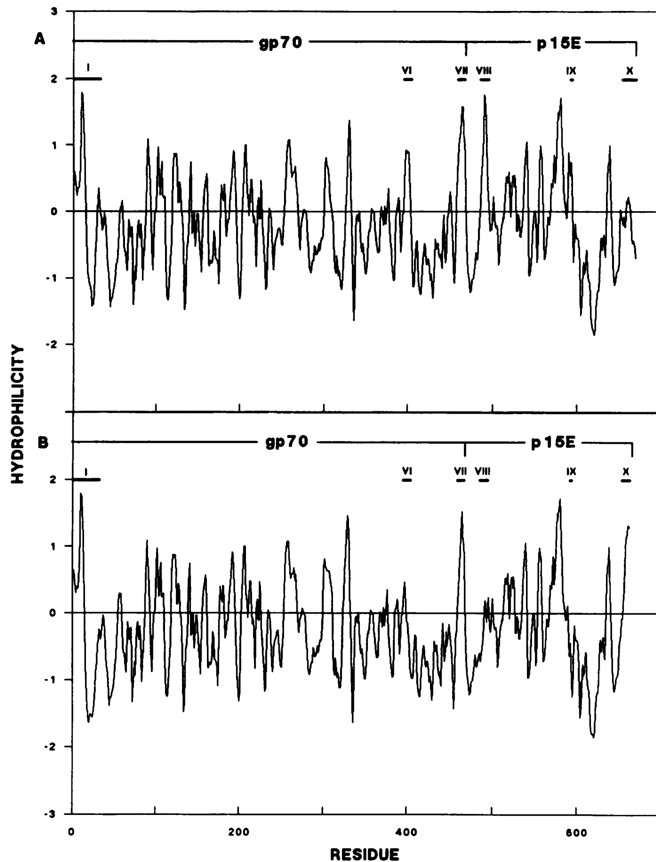

FIG. 3. Comparison of hydrophilicity profiles of *env* gene products of endogenous FeLV (CFE-6) (A) and exogenous FeLV (GA-FeLV-B) (B). The patterns were derived by the Hopp and Woods algorithm (8), using a window size of six residues. Of the 10 variable-homology regions shown in Fig. 2, those which specifically affect the hydrophilicity in this comparison are appropriately marked.

varying in two amino acids from the region in FeLV-C. In contrast, the endogenous p15E was different in all but one amino acid. Moreover, it contained an extended p15E C-terminal sequence of seven amino acids. This type of extension, analogous to that in endogenous xenotropic murine retroviruses (15), had also been detected in another clone of endogenous FeLV DNA (J. Mullins, personal communication).

Because of the increased amino acid sequence homology of the endogenous *env* to FeLV-B relative to other FeLV subgroups, we attempted to compare the hydrophilicity profiles of CFE-6 and GA-FeLV-B to assess the effects of changes in the variable regions. The divergence in hydrophilicity was pronounced in regions VII through X (Fig. 3). Among the other variable regions, I and VI were also affected in hydrophilicity compared with the analogous regions in GA-FeLV-B.

In conclusion, sequence analysis of the endogenous FeLV *env* gene has defined the genetic relationship of this virus to exogenous FeLVs and confirmed the previous speculation that it is more closely related to FeLV subgroup B than to other FeLV subgroups. In the gp70 glycoprotein region, relative to FeLV-B *env*, the endogenous gene contains 30 amino acid substitutions in a total of 465 amino acids. Although most of these deduced amino acid changes are scattered throughout the gene sequence, several cluster in the C-terminal end of the gp70 molecule. The sequence diversity is more pronounced in the transmembrane protein p15E, where three unique regions, VIII, IX, and X, have been identified near the N-terminal, middle, and C-terminal regions, respectively, of the endogenous *env* gene. In addition, our results reveal the presence of an open reading frame in the *env* genes of both of the endogenous provirus loci sequenced and indicate an internal deletion comprising the C-terminal half of gp70 and all of p15E in a naturally truncated endogenous FeLV provirus. Taken together, the data indicate significant sequence variation between endogenous and exogenous *env* genes which may have biological consequences with respect to generation of recombinant FeLVs containing parts or all of endogenous *env* sequences.

We thank B. Bachman for assistance in computer artwork, B. Redmon for preparation of the manuscript, and A. K. Ghosh and E. Mader for critical reading of the manuscript.

D.V.K. was a postdoctoral fellow of the National Institutes of Health (Public Health Service training grant T32-CA09320). This work was supported by Public Health Service grant CA40590 from the National Institutes of Health.

LITERATURE CITED

1. **Berry, B. T., A. K. Ghosh, D. V. Kumar, D. A. Spodick, and P. Roy-Burman.** 1988. Structure and function of endogenous feline leukemia virus long terminal repeats and adjoining regions. J. Virol. **62**:3631–3641.
2. **Dale, R. M. K., B. A. McClure, and J. P. Houchins.** 1985. Rapid single-stranded cloning strategy for producing a sequential series of overlapping clones for use in DNA sequencing: application to sequencing the corn mitochondrial 18 rDNA. Plasmid **13**: 31–40.
3. **Donahue, P. R., E. A. Hoover, G. A. Beltz, N. Riedel, V. M. Hirsch, J. Overbaugh, and J. I. Mullins.** 1988. Strong sequence conservation among horizontally transmissible, minimally pathogenic feline leukemia viruses. J. Virol. **62**:722–731.
4. **Elder, J. H., and J. I. Mullins.** 1983. Nucleotide sequencing of the envelope gene of Gardner-Arnstein feline leukemia virus B reveals unique sequence homologies with a murine mink cell focus-forming virus. J. Virol. **46**:871–880.
5. **Forrest, D., D. Onions, G. Lees, and J. C. Neil.** 1987. Altered structure and expression of c-*myc* and feline T-cell tumours. Virology **158**:194–205.
6. **Hartley, J. W., N. K. Wolford, L. J. Old, and W. P. Rowe.** 1977. A new class of murine leukemia virus associated with the development of spontaneous lymphomas. Proc. Natl. Acad. Sci. USA **74**:789–792.
7. **Hoover, E. A., J. I. Mullins, S. L. Quackenbush, and P. W. Gasper.** 1987. Experimental transmission and pathogenesis of feline AIDS. Blood **70**:1880–1882.
8. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of protein antigenic determinants from amino acid sequences. Proc. Natl. Acad. Sci. USA **78**:3824–3828.
9. **Jarrett, O., M. C. Golder, S. Toth, D. E. Onions, and M. F. Stewart.** 1984. Interaction between feline leukemia virus subgroups in the pathogenesis of erythroid hypoplasia. Int. J. Cancer **34**:283–288.
10. **Levy, L. S., M. B. Gardner, and J. W. Casey.** 1984. Isolation of a feline leukemia provirus containing the oncogene *myc* from a feline lymphosarcoma. Nature (London) **308**:853–856.
11. **Mucenski, M. L., B. A. Taylor, N. G. Copeland, and N. A. Jenkins.** 1987. Characterization of somatically acquired ecotropic and mink cell focus-forming viruses in lymphomas of AKXD recombinant inbred mice. J. Virol. **61**:2929–2933.
12. **Mullins, J. I., D. S. Brody, R. C. Binari, Jr., and S. M. Cotter.** 1984. Viral transduction of *c-myc* gene in naturally occurring feline leukemias. Nature (London) **308**:856–858.
13. **Neil, J. C., D. Hughes, R. McFarlane, N. M. Wilke, D. E. Onions, G. Lees, and O. Jarrett.** 1984. Transduction and rearrangement of the *myc* gene by feline leukemia virus in naturally

occurring T-cell leukemias. Nature (London) **308**:814–820.
14. **Nunberg, J. H., M. E. Williams, and M. A. Innis.** 1984. Nucleotide sequences of the envelope genes of two isolates of feline leukemia virus subgroup B. J. Virol. **49**:629–632.
15. **O'Neill, R. R., C. E. Buckler, T. S. Theodore, M. A. Martin, and R. Repaske.** 1985. Envelope and long terminal repeat sequences of a cloned infectious NZB xenotropic murine leukemia virus. J. Virol. **53**:100–106.
16. **Overbaugh, J., P. R. Donahue, S. L. Quackenbush, E. A. Hoover, and J. I. Mullins.** 1988. Molecular cloning of feline leukemia virus that induces fatal immunodeficiency disease in cats. Science **239**:906–910.
17. **Overbaugh, J., N. Riedel, E. A. Hoover, and J. I. Mullins.** 1988. Transduction of endogenous envelope genes by feline leukemia virus *in vitro*. Nature (London) **332**:731–734.
18. **Riedel, N., E. A. Hoover, R. E. Dornsife, and J. I. Mullins.** 1988. Pathogenic and host range determinants of the feline aplastic anemia retrovirus. Proc. Natl. Acad. Sci. USA **85**:2758–2762.
19. **Riedel, N., E. A. Hoover, P. W. Gasper, M. O. Nicolson, and J. I. Mullins.** 1986. Molecular analysis and pathogenesis of the feline aplastic anemia retrovirus, feline leukemia virus C-SARMA. J. Virol. **60**:242–250.
20. **Sanger, F., A. R. Coulson, B. G. Barrell, A. J. H. Smith, and B. A. Roe.** 1980. Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. **143**:161–178.
21. **Sarma, P. S., and T. Log.** 1973. Subgroup classification of feline leukemia and sarcoma viruses by viral interference and neutralization tests. Virology **54**:160–169.
22. **Sarma, P. S., T. Log, D. Jain, P. R. Hill, and R. J. Huebner.** 1975. Differential host range of viruses of feline leukemia-sarcoma complex. Virology **64**:438–446.
23. **Soe, L. H., B. G. Devi, J. I. Mullins, and P. Roy-Burman.** 1983. Molecular cloning and characterization of endogenous feline leukemia virus sequences from a cat genomic library. J. Virol. **46**:829–840.
24. **Soe, L. H., R. W. Shimizu, J. R. Landolph, and P. Roy-Burman.** 1985. Molecular analysis of several classes of endogenous feline leukemia virus elements. J. Virol. **56**:701–710.
25. **Stewart, M. A., M. Warnock, A. Wheeler, N. Wilke, J. I. Mullins, D. E. Onions, and J. C. Neil.** 1986. Nucleotide sequences of a feline leukemia virus subgroup A envelope gene and long terminal repeat and evidence for the recombinational origin of subgroup B virus. J. Virol. **58**:825–834.
26. **Stoye, J. P., and J. M. Coffin.** 1987. The four classes of endogenous murine leukemia virus: structural relationship and potential for recombination. J. Virol. **61**:2659–2669.