Nucleotide sequence and genomic organization of feline immunodeficiency virus

(retrovirus/lentivirus/phylony)

RANDY L. TALBOTT*, E. ELIZABETH SPARGER†, KATHLEEN M. LOVELACE*, WALTER M. FITCH‡, NIELS C. PEDERSEN§, PAUL A. LUCIV, AND JOHN H. ELDER**

*Department of Molecular Biology, Research Foundation of Scripps Clinic, La Jolla, CA 92037; Department of †Veterinary Medicine and ‡Medical Pathology, University of California at Davis, Davis, CA 95616; and §Department of Molecular Biology, University of Southern California, Los Angeles, CA 90089-1340

Communicated by Frank J. Dixon, May 1, 1989

ABSTRACT An infectious molecular clone of the Petaluma strain of feline immunodeficiency virus (FIV) was isolated from a recombinant bacteriophage library containing genomic DNA prepared from FIV-infected Crandall feline kidney (CRFK) cells. The integrated provirus has a total length of 9472 base pairs. Three long open reading frames corresponding to GAG, POL, and ENV gene coding frames are evident. In addition, an open reading frame overlaps the 3' end of POL, in the region that encodes viral infectivity factor in the primate viruses. Several short open reading frames are present in the intergenic region between POL and ENV and within ENV, which may serve as exons for production of TAT and REV equivalents in FIV. Alignment of the predicted amino acid sequences of the FIV proteins with those of other lentiviruses indicates that FIV did not arise recently from any other characterized lentivirus.

Feline immunodeficiency virus (FIV) is a lentivirus isolated recently from domestic cats suffering from an immune deficiency syndrome (1). FIV causes generalized lymphadenopathy and increased susceptibility to opportunistic infections, eventually culminating in death (2-5). We have characterized the molecular structure of FIV as a prelude to investigations of pathogenesis and development of both vaccine and drug therapies effective against lentivirus infections.

MATERIALS AND METHODS

Cells and Virus. The Petaluma strain of FIV was propagated in Crandall feline kidney cells (CRFK) and purified FIV was prepared from 48-hr harvests of tissue culture supernatants (1). Virus-containing fractions were localized by assaying for the presence of Mg2+-dependent reverse transcriptase activity (1).

Viral RNA and cDNA Preparation. Virus-containing fractions were pooled and viral RNA was prepared as described (6). cDNA was prepared from 500 ng of viral RNA by hybridizing an end-labeled random primer (7). Specific activity (5 × 10⁷ cpm per µg of DNA) was calculated as described (8).

Genomic Libraries. DNA was prepared as described (9) from FIV-infected CRFK cells and partially digested with the restriction enzyme Sau3A to yield an average-sized fragment of 20 kilobases. Fragments were ligated into the BamHI site of the bacteriophage λ vector, EMBL-4 (Stratagene), and six genomic equivalents of DNA were packaged and plated for subsequent screening using viral cDNA (see above). Positive clones were selected and taken through several cycles of purification prior to further analyses.

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: FIV, feline immunodeficiency virus; LTR, long terminal repeat; HIV, human immunodeficiency virus; VIF, viral infectivity factor.

†To whom reprint requests should be addressed.

‡The sequence reported in this paper has been deposited in the GenBank data base (accession no. M25729).
FIG. 2. (Figure continues on the opposite page.)
FIG. 2. Nucleotide sequence and deduced amino acid sequence of the 34TF10 clone of FIV. The N and C termini of the FIV proteins are tentatively assigned, based on alignment to the known N termini of HIV-1 proteins (13, 14). Salient features of the sequence are underlined or indicated by arrows. Features of both LTRs are shown in the 5' LTR for continuity of discussion. GAG, region encoding the core antigens; POL, region encoding protease (pr), reverse transcriptase (rt), and integrase; env, region encoding the major and minor (tm) glycoproteins of the viral envelope; pp, polyuridine tract. In addition to the above gene segments, eight open reading frames (A–H) are evident.

from the original transfection infected both CRFK cells as well as feline peripheral blood lymphocytes (data not shown).

Nucleotide Sequence Analyses. Fig. 2 shows the nucleotide sequence of the 34TF10 isolate of FIV. Salient features of the sequence are indicated with tentative assignments for the boundaries of the viral proteins predicted from alignments with the known N-terminal sequences of human immunodeficiency virus type 1 (HIV-1) proteins (13, 14). A summary diagram of the predicted open reading frames is shown in Fig. 3.

LTRs. The LTRs are 355 base pairs long and are bordered by 2-base-pair inverted repeats (Fig. 2, underlined). Direct sequencing of the viral RNA (12) confirmed that the G residue at base 216 is the 5' end of the viral message (data not shown). The site of polyadenylation is estimated to be around base 9407, predicted by the distance of this site from the polyadenylation signal AATAAA (ref. 15; indicated here in the
frameshifting (28) of transcriptase 3' (14, following aligned between GAG 5746 Biochemistry: Talbott protein encoded the overlaps 3' and protein and protein of with N-terminal protein, 34TF10 overlaps endonuclease of LTR, 5' LTR, bases 269–274, for continuity of discussion). These assignments dictate a U3 region of 215 base pairs, an R region of ≈75 base pairs, and U5 of ≈65 base pairs. In addition to the promoter signal (TATAA) at −26 and the polyadenylation signal at +53, a sequence corresponding to a 9 of 11 match for the NF-κB enhancer element (16) is present in U3 (bases 50–60). The 5' LTR is immediately followed by the 18-base sequence corresponding to the tRNA-Lys primer binding site (bases 358–375; ref. 17), similar to other lentiviruses thus far examined (18–25).

**GAG Gene.** The GAG gene of FIV (bases 627–1976) encodes a predicted polypeptide of 450 amino acids (49,500 Da). The N-terminal protein, p15 (=14,900 Da), begins with the sequence Met-Gly-Asn-Gly, which should be compatible with myristoylation typical of the N-terminal lentivirus GAG protein of HIV-1 (26). Two other proteins, p24 (=24,493 Da) and p10 (=9640 Da), that correspond to the major core protein and nucleic acid binding protein, respectively, are predicted to be encoded by the FIV GAG gene.

**POL Gene.** The POL gene (bases 1868–5239) overlaps the 3' end of the GAG gene. The POL polypeptide is presumably synthesized as a GAG-POL polypeptide by ribosomal frameshifting, similar to HIV-1 (27). Two consensus sequences for frameshifting (28) are present in the overlapping region between GAG and POL (underlined), and the second precedes immediately the predicted N terminus of the protease protein (or, ref. 29). Alignment with the known N termini of HIV reverse transcriptase and endonuclease (14) predicts a protease of 13,493 Da and a reverse transcriptase of 61,490 Da.

In the primate lentiviruses, the endonuclease protein immediately follows reverse transcriptase in the polypeptide (14, 18–23). However, in the nonprimate lentiviruses including EIAV (24) and visna (25) as well as FIV (Fig. 2; denoted by the symbol -), bases 4013–4405 there are ≈400 base pairs intervening between reverse transcriptase and endonuclease when aligned with the primate POL genes (18–23). This region could thus encode a protein product of ≈14,600 Da. The 3' end of the polymerase gene is predicted to encode an endonuclease of 30,690 Da.

**Viral Infectivity Factor (VIF)?** A third open reading frame overlaps the 3' end of the polymerase gene (Fig. 2, base 5199) and is preceded by a possible splice acceptor site (underlined). The location suggests that this region corresponds to the VIF of the primate viruses (18–23), although alignment with both primate lentiviruses and visna virus (25) yielded no significant similarity.

**Intergenic Region.** The region 3' to VIF has the potential to encode three short overlapping reading frames (Figs. 2 and 3, A, B, and C) not found in other characterized lentiviruses (18–25, 30). This finding implies that either FIV differs in splicing strategy from the other lentiviruses or that this region of FIV has no coding capacity. Another possibility is that the 34TF10 isolate contains changes in this region that close the reading frame (see below). Clearly, if the latter is true, the protein encoded by this region is not critical for viral replication or infectivity.

**ENV Gene.** The FIV envelope gene (Figs. 2 and 3, ENV) overlaps the 3' end of open reading frame C. This designation is shown in parentheses because there are ≈750 bases of coding capacity prior to the presumed hydrophobic leader sequence (encoded by bases 6710–6797, underlined) for ENV. It has recently been demonstrated that a portion of the region preceding the ENV leader of visna virus (25) is the first coding exon of a gene product (L; ref. 31) of unknown function. Possibly, a similar product is generated by FIV, although no sequence similarity is evident.

The amino acid sequence following the presumed ENV leader sequence contains 17 potential sites for N-linked glycosylation (brackets) prior to the predicted processing site for the minor glycoprotein at base 8096 (Arg-Arg-Lys-Arg, underlined). The minor glycoprotein (Fig. 2, tm) contains four potential N-linked glycosylation sites (brackets) and a hydrophobic membrane spanning region (bases 8619–8690, underlined). Two additional short open reading frames (E and F) are evident near the beginning of the minor glycoprotein.

No continuous open reading frame is present after the end of the ENV coding region at base 8832 of the 34TF10 isolate of FIV. However, this isolate differs from other FIVs we have tested in that it infects CRFK cells productively. Since this location encodes a negative factor in the primate lentiviruses (18–23), which may down-regulate virus expression, a potential protein product overlapping with the end of ENV (Figs. 2 and 3, G?) is indicated. The potential coding sequence overlaps the U3 region of the 3' LTR to base 9165 but contains two stop codons (!) in the 34TF10 isolate. It is possible that mutations in this region are responsible for the ability of this isolate to infect CRFK cells. Another short open reading frame (H) begins at base 8957 and terminates at base 9890, short of the polyuridine tract (underlined) that precedes the beginning of the 3' LTR.

**Phylogenetic Analysis.** FIV diverges from the other lentiviruses throughout the genome. However, homologies within reverse transcriptase served to compare the relative relatedness of FIV to the other lentiviruses. Fig. 4 shows the most

![Fig. 3. Graphic representation of the open reading frames of the 34TF10 clone of FIV. In addition to the coding regions of GAG, POL, and ENV, the virus contains eight short open reading frames (A–H).](image)

![Fig. 4. Phylogenic tree of lentivirus reverse transcriptases.](image)
parsimonious phylogenetic tree, prepared at the nucleotide level using the first 420 amino acids of reverse transcriptase. This analysis indicates that FIV diverged early relative to the initial split between the primate and nonprimate lentiviruses. However, there is clearly an initial split between the primate and nonprimate lentiviruses. The percentage similarity of FIV to the other lentiviruses ranges from 41% to 45% within reverse transcriptase.

**DISCUSSION**

The results indicate that FIV shares many features with other lentiviruses. However, there is no molecular evidence to suggest that FIV came from another species recently. Clearly, FIV has the coding capacity for several possible protein species other than those encoded by GAG, POL, and ENV. It may be that the 34TF10 isolate contains mutations, particularly in the intergenic region between POL and ENV, resulting in three short open reading frames rather than one continuous reading frame. Preliminary analysis of another FIV isolate supports this contention in that the nucleotide sequence of the second isolate varies so as to encode one continuous gene product in the intergenic region between POL and ENV (unpublished data). Such mutations clearly do not interbreed with viral replication, since the 34TF10 isolate is fully infectious. FIV will serve as a valuable small animal model for the study of lentivirus infection, etiology, and treatment.

The authors wish to thank the HIV data base for additional phylogenetic analyses. The excellent technical assistance of Karen Shaw is acknowledged. We thank Hoffmann-La Roche for donation of interleukin 2. This research was supported by grants from the National Institute of Allergy and Infectious Diseases to J.H.E. (RO1-AI25825) and N.C.P. (RO1-AI25802) and from the National Science Foundation to W.M.F. (BSR 8796183). E.E.S. is a recipient of a National Research Service Award from the National Institutes of Health and P.A.L. is a recipient of an Investigative Award from the State of California University-wide Task Force on AIDS.