Comparison of the complete sequence of feline spumavirus with those of the primate spumaviruses reveals a shorter gag gene

C. R. Helps and D. A. Harbour

Department of Clinical Veterinary Science, Division of Molecular and Cellular Biology, University of Bristol, Langford House, Langford, Bristol BS18 7DU, UK

The complete nucleotide sequence of the provirus of feline foamy virus (FeFV), strain F-17, was determined, and compared to the available data for human and simian spumaviruses. In addition to the usual retroviral gag, pol and env genes, two open reading frames are present between the env gene and the 3'-LTR, as in the simian spumaviruses, the first being the putative transactivator. The gag gene is predicted to encode a precursor protein of only 53 kDa compared to 70 kDa for simian spumaviruses and a doublet of 70/74 kDa for human spumavirus. The gag gene contains conserved splice acceptor and donor sites suggesting that, like human foamy virus, FeFV expresses its pol gene using a spliced mRNA. The pol and env genes showed greater sequence similarity to their counterparts in the primate spumaviruses than the gag gene and additional open reading frames.

Introduction

Spumavirus is one genus of the family Retroviridae, but unlike other genera within the family, spumaviruses have received little attention until recently. Spumaviruses have been isolated from several mammalian species, including domestic (Kasza et al., 1969; Riggs et al., 1969) and wild (Lieber, 1975) cats, hamsters (Hruska & Takemoto, 1975), sea-lions (Kennedy-Stoskopf et al., 1986), cattle (Malmquist et al., 1969), sheep (Flanagan, 1992), man (Achong et al., 1971) and several non-human primates (Hooks & Gibbs, 1975). The feline isolates were called feline syncytium-forming virus (FeSFV), but are now known as feline foamy virus (FeFV) to comply with the recently adopted nomenclature. FeFV can readily be isolated from healthy cats, and although infection has been statistically associated with chronic progressive polyarthritis in the cat (Pedersen et al., 1980), the disease has not been reproduced experimentally, and FeFV is generally considered to be apathogenic.

Spumaviruses have a number of advantages over other retroviruses as potential integrating gene transfer vectors, including their larger genome size and wide host-range. Moreover, they do not require polycations for efficient transduction, which may be an advantage in vivo, and since the transactivator is required for high-level transcription from the viral LTR, inactivation of the transactivator gene should prevent activation of downstream cellular genes by promoter insertion (Russell & Miller, 1996).

As a preliminary to developing FeFV as a gene transfer vector, we have determined the complete nucleotide sequence and genomic organization of the proviral genome of FeFV. No attempt was made to create a complete molecular clone, although this work is now in progress.

Methods

- **Viral RNA and total DNA extraction.** FeFV strain F-17 (Riggs et al., 1969) was purchased from the American Type Culture Collection (ATCC VR-889) at passage 8 in feline cells (six times in feline thymocytes and once each in feline embryo fibroblasts and Crandell feline kidney cells). The virus was propagated in subconfluent WFE cells (Harbour et al., 1991) grown in Eagle’s MEM supplemented with 1% (v/v) non-essential amino acids, 5 ml/l 10% (w/v) lactalbumin hydrolysate solution, 5% (v/v) foetal calf serum and 50 µg/ml gentamicin. Virions were concentrated, from the supernatant of cultures showing > 75% CPE, by centrifugation at 20000 r.p.m. in a Kontron TFA 20.250 rotor for 16 h, and purified through two 30–60% (w/v) sucrose density gradients at 40000 r.p.m. for 1 h in a Kontron TST 55.5 rotor. Virion RNA was extracted using a QIAamp HCV kit (Qiagen) according to the
Results and Discussion

Nucleotide sequence and open reading frames

The genomic organisation is shown in Fig. 1, and the complete nucleotide sequence is shown in Fig. 2. The genome is 10456 nucleotides long [excluding the poly(A) tail], the provirus is 11657 nucleotides and the LTRs are 1336 nucleotides in length. The 5'-LTR is 10456 nucleotides long [excluding the poly(A) tail], the provirus is 11657 nucleotides and the LTRs are 1336 nucleotides in length. The 5'-LTR is bordered by a polypurine tract at position 10303–10321 and contains a putative polyadenylation signal

env

and TATA transcriptional initiation signal at position 1026–1029. The viral protease of FeFV appears to be located in the N-terminal region of the pol gene, but the highly conserved protease sequence Asp-Ser-Gly-Ala Thr found in simian and human spumaviruses is altered to Asp-Ser-Gln-Ala-Asp in the feline virus.

Comparison of FeFV with primate spumaviruses

The complete proviral nucleotide sequence of FeFV, strain F-17, was analysed and compared to that of primate spumaviruses. The genome is the smallest of the spumaviruses sequenced to date, due mainly to a shorter gag gene. Like simian spumavirus types 1 and 3, FeFV has two additional ORFs apart from the usual retrovirus gag, pol and env genes, the first of which encodes a putative transactivator protein. The function of the products of the other one or two ORFs is unknown.

The predicted size of the Gag precursor protein of FeFV strain F-17 is only 53 kDa compared to 70 kDa for the two simian spumaviruses and 70/74 kDa for the human spumavirus. Since this virus had been isolated in 1969 in the USA, the gag gene from a recent (1993) British isolate (A491) was amplified by PCR and shown to be the same size (C. R. Helps & D. A. Harbour, data not shown). This is similar to the size of the Gag precursor protein reported for FeFV isolate FUV (Bodem et al., 1990), so a smaller gag gene appears to be a consistent feature of FeFV isolates. A comparison of the amino acid sequence of FeFV and simian spumavirus Gag proteins show that they have only 38% identity. However, a comparison of their hydrophilicity plots (Fig. 3) indicates that their overall structure appears very similar despite their difference in size. Like human foamy virus (Morozov et al., 1996), there are no zinc-finger motifs typically found in other retroviral nucleocapsid proteins, but there are glycine/arginine-rich regions (GR boxes) in the carboxyl third (residues 360–366, 426–432 and 468–475). At least one GR box (RPSRGRGRGQ) in human foamy virus (HFV) has been shown to be essential for nucleic acid binding and virus replication (Yu et al., 1996b), although this sequence is not perfectly conserved in FeFV (GPGRGGRRGQ).

Comparisons of the nucleotide sequence of FeFV strain F-17 with the human and simian spumaviruses revealed that the pol gene showed 64% identity and the env gene 54–55% identity, whereas the gag gene showed only 46–47% identity. The highest level of conservation was between the pol genes of FeFV and human spumavirus. The putative transactivator (ORF1) showed only 39–44% identity to its counterparts in the primate and human viruses, and there was little similarity between ORF2 and the other two 3′-ORFs of the human virus (41% similarity to ORF2 and 38% similarity to ORF3) or the most 3′-ORF of the simian viruses. Like HFV bel 1 (He et al.,
FeFV ORF1 does not contain sequences homologous to known transcriptional activation or DNA-binding motifs. The most 3′-ORF (ORF2) of FeFV terminates in the U3 region of the 3′-LTR, which is in agreement with the findings for both human and simian spumaviruses (Renne et al., 1992).

Analysis of the spumaviruses at the molecular level is still rather limited. One isolate of human spumavirus (Maurer & Flugel, 1988), an isolate from chimpanzees (Herchenroder et al., 1994) and simian spumaviruses types 1 (Kupiec et al., 1991) and 3 (Renne et al., 1992) have been molecularly cloned and the sequence and genomic organization determined. Partial sequences are available for a number of other spumaviruses (Bieniasz et al., 1995). The human spumavirus provirus has a genome 12–2 kb long (Rethwilm et al., 1987). It encodes the usual gag, pol and env genes common to all replication-competent retroviruses, and encodes a further three ORFs at the 3′ end of the genome (Renne et al., 1992; Maurer & Flugel, 1988). The first of these has been shown to encode the transactivator protein (Lee et al., 1992; Keller et al., 1992; Venkatesh et al., 1991; Rethwilm et al., 1990). Simian spumaviruses types 1 and 3 encode only two ORFs at the 3′ end of the genome (Renne et al., 1992) but in each case the first of these encodes a transactivator. The transactivator of simian spumavirus type 1 has been shown to transactivate human immunodeficiency virus and simian immunodeficiency virus but not human spumavirus (Mergia et al., 1992).

As expected, whereas the env gene of lentiviruses shows the most variation between isolates, in the spumaviruses sequenced to date, the env gene is relatively highly conserved, and the gag gene shows greater variability (Renne et al., 1992; Herchenroder et al., 1994; this study).

The gag and pol genes of the primate spumaviruses overlap by 22–92 bp, and are in +1 frameshift (Renne et al., 1992), but those of FeFV do not overlap, and are in frame. Analysis of the nucleotide sequence between the end of the gag gene and the beginning of the pol gene showed no classical heptanucleotide sequence (ACAAATT) associated with ribosomal frame-shifting in most other retroviruses, which are predicted or known to express pol genes as Gag–Pol or Gag–Pro–Pol polyproteins by −1 ribosomal frameshifting (Varmus, 1988) or, in the case of murine leukaemia virus, by suppression of the gag amber termination codon (Yoshinaka et al., 1985). Bodem et al. (1996) showed that, like HFV (Yu et al., 1996a; Lochelt & Flugel, 1996), FeFV isolate FUV expresses its pol gene as a Pro–Pol polyprotein using a spliced mRNA. The splice donor and acceptor sites are perfectly conserved in isolate F-17, and it is assumed that F-17 uses a similar splicing mechanism.

The envelope proteins of feline and simian spumaviruses show 43% amino acid identity and both contain three stretches of hydrophobic amino acids (66–89, 575–589 and 936–972 in FeFV) which are conserved in their distribution. The first hydrophobic region is preceded by a large hydrophilic domain of 65 amino acids which extends to the N terminus of the protein and is presumed to act as a signal for the export of SU domain to the cell surface. This large hydrophilic domain is also found in the simian and human spumavirus envelope proteins (Renne et al., 1992). The second hydrophobic region is preceded by a string of basic amino acids (Lys-Arg-Gln-Arg-Arg) at position 557–561, which is predicted to be the proteolytic cleavage site of the Env precursor. The third hydrophobic region is 37 amino acids long and is the putative transmembrane domain of the envelope protein. Both envelope proteins contain 16 potential sites for N-glycosylation (Asn-X-Ser/Thr); 11 of these are located in the SU domain of FeSFV and 5 in the transmembrane domain. This differs slightly from the distribution in SFV-3 where 13 sites are found in the SU domain and only 3 in the transmembrane domain.

Analysis of the 5′-LTR nucleotide sequence of FeFV revealed several cellular transcription factor binding sites. In particular, the nucleotide sequence CTGACTAA at position 462–469 is a potential API-CS1 transactivator binding site and three API-CS3 binding sites were found at positions 314–320, 463–469 and 568–574. These findings agree with those of Maurer et al. (1988) who showed that there were three API binding sites in the U3 region of human spumavirus.

Phylogenetic trees were constructed for gag, and partial pol, env and LTR sequences using the method of Thompson et al. (1994) to align sequences, and the neighbour-joining method of Saitou & Nei (1987) to construct the trees. Using the LTR R/U5 sequences (Fig. 4) the two feline isolates, F-17 and FUV, were closely related to each other and clearly dissimilar to the primate isolates. The human and chimpanzee isolates were closely related, with the isolates from other primate species being more divergent (Bieniasz et al., 1995). The groupings

![Fig. 1. Genomic organization of FeFV. The upper panel shows the LTRs, gag, pol and env genes together with the two open reading frames ORF1 and ORF2. Three overlapping cDNA clones are shown in the lower panel. Numbers represent distance along the genome in base pairs.](image-url)
Fig. 2. For legend see page 2561.
Feline spumavirus sequence

```
1370 1380 1390 1400 1410 1420 1430 1440
  * * * * * * * *
  TGGAGTGAATATAAATTCAGCTAGGAAGAATAATCTCCAGGGCCTTACCTTATGAGAGATGGCTGAGTTAA
gag MetAlaArgGluLeu

1450 1460 1470 1480 1490 1500 1510 1520
  * * * * * * * *
  ATCCCTCCATACGCACTGTAATTAAATATTAGTTTACACCCCTATCCAGGAACTGAGAGATTGTGCTGCA
AsnProLeuGlnLeuGlnLeuYrIleAsnAsnGlyLeuGlnProAsnProGlyHisGlyAspValIleAlaValArg

1530 1540 1550 1560 1570 1580 1590 1600
  * * * * * * * *
  TTTACAGGGAGACCGCCGCTGGGCTGAGTATGAGTGGACAGTCAAGTTGAGCAGACGACCCACCTTT
PheThrGlyGlyProTrpGlyProGlyAspArgTrpThrArgValIleArgLeuGlnAspThrGlyGlyGlnProLeu

1610 1620 1630 1640 1650 1660 1670 1680
  * * * * * * * *
  ACAGTTCCCTGAATGGGTTGAGCCGACCCGCTCAGTAAGCTAATATCTGAGAGACAGACGACCCACCTTTA
GlnValProGlyGlyLeuGluProGlyIleIleAlaAsnArgGluAspIleLeuIleAlaGlyProTrpAsnLeu

1690 1700 1710 1720 1730 1740 1750 1760
  * * * * * * * *
  TAAAGAATCGCTTTCCTGGATTAGGCTGCAAGCGACGACGACGACGCCGCTTCTGTTGAGTACGAGATTTACCT
IleArgThrAlaPheLeuAspLeuGluProAlaArgGlyProGluArgHisGlyProPheGlyAspArgLeuGlnPro

1770 1780 1790 1800 1810 1820 1830 1840
  * * * * * * * *
  GGAGATGTTTATCTGAGGGTCTTACATCTACTGAGTAAAGCTAATGTATCCAGGACGACGACCCACCTTTAG
GlyAspGlyLeuSerGlyLeuGluProGlyIleIleThrAspGluMetGlnAlaValGlyThrIleGlyAlaAlaArg

1850 1860 1870 1880 1890 1900 1910 1920
  * * * * * * * *
  AATAAGGAGATAGTGTCGACGACCTACGATAGAACTGACGACGACGACGACGACCCACCTTTAG
AsnGluValArgLeuArgLeuGluAlaLeuGlnArgLeuGlnValGlyGlyValGlyArgProIleProGlyAlaIle

  * * * * * * * *
  TACACCAACACGATATAGGGCCCTGTATACAGCTATACATCTACTGAGTAAAGCTAATGTATCCAGGACGACGAC
LeuGlnProGlnProValIleGlyProValIleProIleAsnHisLeuArgSerValIleGlyAsnThrProAsnPro

2010 2020 2030 2040 2050 2060 2070 2080
  * * * * * * * *
  CGAGATGCTGACCTCTAAGGCTGAGATCCACAGGCCTATGAGGTGATTTCCCCTATAGGGACAAACCCCATAT
ArgAspValAlaLeuTrpLeuGlyArgSerThrAlaIleGluGluGlyValPheProIleValGspInIleThrArgMet

2090 2100 2110 2120 2130 2140 2150 2160
  * * * * * * * *
  GAGGAGTATGTTATGAGGGTCTTACATCTACTGAGTAAAGCTAATGTATCCAGGACGACGACGACCCACCTTTAG
ArgValValAlaLeuValAlaSerHisProGlyLeuThrLeuGluGluGluGlySerTrpAsnAlaAla

2170 2180 2190 2200 2210 2220 2230 2240
  * * * * * * * *
  TATGACCTTATGGAGGCAAGCGTATGAGGGGCTGACGCACTGATGAGGAGATTTACCCTATGAGGAATTTACAT
IleSerAlaLeuTrpArgGlyAlaHisGlyAlaAlaAlaHisGluLeuAlaGluValSerAspIleAsnGlyLys

2250 2260 2270 2280 2290 2300 2310 2320
  * * * * * * * *
  GAGGAGTATGCGCAGCTACATCTACTGAGTAAAGCTAATGTATCCAGGACGACGACGACGACCCACCTTTAG
GluGlyIleGluThrAlaPheAsnLeuGlyMetGlnPheThrGlyAspArgAsnTrpSerLeuValTrpGlyIleThrArgThr

2330 2340 2350 2360 2370 2380 2390 2400
  * * * * * * * *
  TCCCTACGCCGAGAATGCTACGCTCATGCAAATGCAATGCAAATGCAATGCAAATGCAATGCAAATGCAATGCAA
LeuLeuProGlyGlnAlaLeuThrAsnAlaGlnSerGlnProAspLeuMetGlyAspIleGlnGlnArgAlaGlu

2410 2420 2430 2440 2450 2460 2470 2480
  * * * * * * * *
  ATTGTACCGTATTAATTAATTTATACCTATGCTCGCTGCTATATATATACCGCGACGCGATTTAGCAGCAGCGGCCAAA
AsnPheProArgValIleAsnAsnLeuThrThrMetLeuGlyLeuAsnIleHisGlyGlnSerIleArgProArgValGln
```
CTCAGTGCGAGAAGGAAGAATTACATTTGGAATCAGTGCCTCTACCATCCCTCAAGCCAGTGGAAAGGTGAAAGG
AlaGlnTrpAlaLysGluLysAsnIleGlnLeuGluPheSerAlaProYrHisProLeuSerSerGlyLysValGluArg
5850 5860 5870 5880 5890 5900 5910 5920

AAAAACAGGGAAATTTAAGAAGCTTACATTCTTAGGTCAGCAGCAGTACATACATGCTACATGCTACATGCTACAT
LysAsnSerGluLysLysLeuLeuThrLysLeuLeuValGlyArgProLeuLysTrpPheAsnLeuIleSerSerVal
5930 5940 5950 5960 5970 5980 5990 6000

GCAAACCTGCCCTAAATACCTCACTGTTGCTGACACCAAGATAGTCGTACATCAAGCAGCTCTTGAGAATG
GlnLeuAlaLeuAsnAsnThrHisValValSerThrLysThrThrProHisGlnLeuMetPheGlyIleAspCysAsn
6010 6020 6030 6040 6050 6060 6070 6080

TACATTGTGAAATACAGGTAACCTTGGACTGCGAAGAAGAAGAAATCTGCTTGGAAATATTTAACTGAGCTTTGATG
LeuProPheAlaAsnLysAspThrLeuGluGluGluLeuAlaLeuLeuGlnGluIleArgGluSerLeu
6090 6100 6110 6120 6130 6140 6150 6160

CAACACCGCTGGCACCACCCACCTCCTCCTTGTCGCTACATGCGACAGGCTGATCAGGTCGACAGGTCGAC
GlnHisProValGlnProSerThrSerGlyLysGlyTrpSerProGlyValGlnGluValGluArgValGluArgSer
6170 6180 6190 6200 6210 6220 6230 6240

GTCACAACTAAGCCCTAAATGCGGACGGCTAACAAAGCTTTGAAATATTTAACTGAGCTTTGATG
SerGlnLeuArgProLysTrpArgGluProThrLysLeuValLeuIleLeuLeuAsnProArgThrValIleGluAsp
6250 6260 6270 6280 6290 6300 6310 6320

AYCTAGGCCACGGAAATCTGTTGATATGCAATTTAAAAACCAACAGACCATGCAATAAGCCACCAAGACGATG
HisLeuGlyGlnArgLysSerValSerIleAspAsnLeuLysProThrAlaHisGlnHisAsnGlyThrArgThrCysAsp
6330 6340 6350 6360 6370 6380 6390 6400

GACCTGGAAGGATAGTCTGATTGTTCTCATACAAACTCCAGAACTGCTGGAATAGATGTCTTACTCTGTG
AspLeuGluGlyMetAspGlyGluCysSerGlnThrThrGluThrThrSerValAspSerSer
6410 6420 6430 6440 6450 6460 6470 6480

ACATACCCTGAGATATCATTCACCATCAGAGAGAAGCAGTTGCTAACTCTCTTACTACTCTTCTCTTCCTAGTGT
AspIleProGluAspIleHisSerValProGluValProLeuAsnMetArgMetArgGluArgGluArgCysThrLeuCys
6490 6500 6510 6520 6530 6540 6550 6560

GCTACTCTACTGATATGTTTTGACTACTTTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
AlaThrSerThrArgIleMetPheProIleLeuPheProLeuPheLeuCysPheSerIleValThrSerThrIleIleSer
6570 6580 6590 6600 6610 6620 6630 6640

TACTCTGATTACATGAAAGAAAGCAATACACATCCTCTGCACGGACGACGGATAGATACATACACACCCAGTGAC
ThrMetGlyGlyAsnThrSerSerSerArgArgArgAspIleGlnTyrHisLeuProValGluValAsnIle
6650 6660 6670 6680 6690 6700 6710 6720

CAATGGCAAGAGATTACACCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
ThrMetGlyGlyAsnThrSerSerSerArgArgArgAspIleGlnTyrHisLeuProValGluValAsnIle
6730 6740 6750 6760 6770 6780 6790 6800

TCAGCGACCCAGACGCTTCTCTGCTCACCACAAACCTACACTTCTACAGAAAGAACACTTCTGAGTCTTCTCTC
SerGlyIleProGluGluLeuPheAspIleProGluProLysIleLeuHisGluArgThrLeuGlyLeuSerGln
6810 6820 6830 6840 6850 6860 6870 6880
AGTGATCTTTACTGGATCGCACTCATATATACAGCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAAG
Val1Leu1Leu1Leu1Asp1Ser1Thr1Thr1Glu1Gly1His1Tyr1Lys1Glu1Val1Ser1Val1Ser1Thr1Ile1Asn1
6890 6900 6910 6920 6930 6940 6950 6960
*  *  *  *  *  *  *  *
AAGAAATGGCACTAAAATAAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATACATATCTGCTACAATT
Glu1Glu1Met1Glu1Len1Leu1Lys1Thr1Val1Leu1Pro1Phe1Asp1Leu1Pro1Thr1Lys1Asp1Pro1Leu1Thr1Gln1Glu1Lys1
6970 6980 6990 7000 7010 7020 7030 7040
*  *  *  *  *  *  *  *
GAGAAGAGCTCTTCAACTAATTTGGACATTATATAGTAATGATGAAATCCCTAAGAAAAACGGCATTTGATGTTTCCCAAAG
Glu1Lys1Arg1Glu1Asp1Gln1His1Phe1Gly1His1Cys1Thr1Val1Leu1Glu1Thr1Gly1Ser1Pro1Arg1Gly1Trp1Pro1Phe1Asp1Asn1
7050 7060 7070 7080 7090 7100 7110 7120
*  *  *  *  *  *  *  *
AAATACAGAATCCCTTTACTCTGCTGATATACGTTAAGGCGCTCGATATAGAAATCTCCTGTTTGGATGATGTTTCCCAAAG
Glu1Lys1Asp1Gln1Pro1Glu1Pro1Val1Leu1Ser1Asn1Trp1Ser1Ser1Pro1Gly1Asp1Ala1Arg1Glu1Gly1Ser1Phe1Glu1Val1Pro1Lys1
7210 7220 7230 7240 7250 7260 7270 7280
*  *  *  *  *  *  *  *
GAATTTAAAAGAAATCTCCTCACTAAATCTGTTGCTCACATAATACGTTAAGGCGCTCGATATAGAAATCTCCTGTTTGGATGATGTTTCCCAAAG
Glu1Phe1Lys1Asp1Glu1Ala1Thr1His1Gly1Ile1Phe1Cys1Ser1Asp1Gln1Len1Leu1Glu1Thr1Gly1Trp1Glu1Pro1Arg1Thr1Leu1Pro1Ser1
7290 7300 7310 7320 7330 7340 7350 7360
*  *  *  *  *  *  *  *
TCAGACTTGGAACCTACTAAAACCTTTAAATTGAAGATATATATGAAATCTCCTGTTTGGATGATGTTTCCCAAAG
Glu1Thr1Leu1Glu1Len1Leu1Ala1Thr1Phe1Glu1Leu1Ala1Ser1Thr1Pro1Leu1Met1Arg1Ile1Leu1Ser1Arg1Glu1Gly1Asn1Glu1Leu1
7370 7380 7390 7400 7410 7420 7430 7440
*  *  *  *  *  *  *  *
CATCTGTTGCTACTTTGCTCTAAGGAACAAAACTCCTCTCAGAGATTGTTCCATATAGATATGTTAGTACACAACTA
Ser1Ser1Leu1Ala1Pro1Thr1Ser1Leu1Ser1Ser1Phe1Gly1Len1Lys1Leu1Leu1Pro1Arg1Ser1Phe1Val1Asp1Ser1Cys1Asn1Ile1
7450 7460 7470 7480 7490 7500 7510 7520
*  *  *  *  *  *  *  *
CCATAAGGAGCTTCACTTTAAACAGAACTTTATGTTGACATTTTCTCTGTTGCGAATTCTACAGACAAA
Pro1Lys1Ala1Val1Leu1Leu1Asn1Arg1Thr1Trp1Pro1Pro1Phe1Ser1Leu1Glu1Trp1Glu1Asp1Phe1Ile1Pro1Gln1Glu1Thr1Asn1
7530 7540 7550 7560 7570 7580 7590 7600
*  *  *  *  *  *  *  *
TATCAAGAATCTCCTGTTGCAATGCTAACCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATACAT
Ile1Thr1Glu1His1Ala1Ser1Cys1Lys1Phe1Asn1Arg1Thr1Gln1Thr1Ser1His1Pro1Thr1Ala1Cys1Thr1Phe1Trp1Arg1Gln1
7610 7620 7630 7640 7650 7660 7670 7680
*  *  *  *  *  *  *  *
ATTAGGAAGATTTGGATGCAATGCTAACCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATACAT
Tyr1Leu1Glu1Asp1Ser1Ile1Lys1Val1Asp1Ser1Ala1Arg1Cys1Tyr1Ser1Pro1Ala1Thr1Gly1Val1Glu1Asn1
7690 7700 7710 7720 7730 7740 7750 7760
*  *  *  *  *  *  *  *
AGGAGAATTTGGATGCAATGCTAACCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATACAT
Arg1Glu1Asp1Phe1Gly1Trp1Gln1Ala1Tyr1Asn1Glu1Asp1Phe1Ser1Pro1Ser1Val1Pro1Cys1Ile1Lys1Glu1Ile1Pro1Ile1Lys1
7770 7780 7790 7800 7810 7820 7830 7840
*  *  *  *  *  *  *  *
GAATTAAGATTTGGATGCAATGCTAACCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATAC
Asn1Tyr1Leu1Val1Ser1Ser1Val1Leu1Lys1Ile1Asn1Val1Asp1Ala1Lys1Val1Ile1Lys1Glu1Val1Ile1Asp1Glu1
7850 7860 7870 7880 7890 7900 7910 7920
*  *  *  *  *  *  *  *
TGGAAATTTGGATGCAATGCTAACCAGCCATATTAAATAACACGGATAGCAATACATATCTGCTACAATTAAATAACACGGATAGCAATACAT
Leu1Glu1Asp1Leu1Phe1Ser1Thr1Lys1His1Val1Leu1Pro1Glu1Asp1Thr1Pro1Lys1Pro1Ser1Leu1Glu1Asp1Thr1Trp1Pro1Lys1
7930 7940 7950 7960 7970 7980 7990 8000
*  *  *  *  *  *  *  *
Feline spumavirus sequence

```
GAAACACAAATACAAACAAAAACTATCTTGTAGGAAGGAAACAAAAAGACAAAGAGAGGTTGCCAGAAGAGAAA
GlulysGlnAsnLysGlulysThrSerGlySerLysAsnLysArgGlnArgGlnArgSerValSerThrGluAsm
8010 8020 8030 8040 8050 8060 8070 8080
```

```
CTTAAGAGAAATGCAAAGGGAGCTATGCTCAGGCAATAGCTCATCATTAGCAGACTGGTATGCTCAAGACA
CLeuArgArgMetGlnValAlaGlyLeuGlyLeuAlaAsnAlaIleThrThrValAlaLysIleSerAspLeuAsnAsp
8090 8100 8110 8120 8130 8140 8150 8160
```

```
AGAAMTAAGCCAGGAGTACATTGGCTTGGAGATCATGCTATGCTCAAGACACTGGTATGCTCAAGACAT
GlnLysLeuAlaLysGlyValHisLeuLeuArgAspSerValValThrLeuMetGlulysAlaAsnLeuAspAspIleValSer
8170 8180 8190 8200 8210 8220 8230 8240
```

```
CATAGAGGGATCAAAAAAGAAGATCACTACAATACACACTATCTCTGAATAGCTATGGCAAGAAGAAACTG
LeuGlyGlyIleGlnIleGluHisIleHisAsnIleSerLeuLeuLeuLeuAsnArgIleAsp
8250 8260 8270 8280 8290 8300 8310 8320
```

```
CTGGAGGTATATATATGCTACTAGATAGATGCTGCCAGGAGATAGCTATGCTCAAGACA
TrpArgArgPheIleAsnSerGluTrpIleGlulysLeuGlyValSerAspAsnIleMetLysValIleArgLysh
8340 8350 8360 8370 8380 8390 8400
```

```
CAAGGAGTTATCCTAAGCTAAAACAAACAGAATCTAAAACACTCTACTGCTATGGGAATATATATTATGAG
AlaArgCysIleProThrAsnValLysGlnThrArgAsnLeuAsnSerThrThrAlaTrpIleGluIleLeuLeuTyrGlu
8410 8420 8430 8440 8450 8460 8470 8480
```

```
AATTAAACTCTCTACCTACATACAAAGAGATGGAATATAAAGATCTATAGCTACTCTGCTATGGCAAGATTT
IleIleIleProThrIleTyrGlnAsnTrpAsnIleLysGluLeuArgAsnAsnAlaGlyTyrLeu
8490 8500 8510 8520 8530 8540 8550 8560
```

```
GCTTAAGGTTGGGTTCCAACACAAATCTTTAAAGATATATTTAAAATACAGAATGGTAACAAATATATATATAT
SerLysValTrpIleGlnGlnProGluValLeuGluAsnGluCysGlyThrAsnIleTyrLeuHisMetGlulys
8570 8580 8590 8600 8610 8620 8630 8640
```

```
GTTTGGCCAGCAGTCTATATATATATGGAGAGTTAGATAGATGGAATATATATTAGGGGAAAACTGTTGTCAGCTGCCCA
CysValAspGlnAspTyrIleIleCysGluGluValIleGluPheProProCysGlyAsnGlyThrGlySerAspCysPro
8650 8660 8670 8680 8690 8700 8710 8720
```

```
GCTTAACTTTAACCCTTCCAGAGATATCTTGGAATATGACCCCTGAGAATGAGATTGCTATTGTTTTGATCAAGTAC
ValLeuThrLysProLeuThrLeuGluLeuValIleGluLeuAlaProValValAspLeuAspSerThrSerThr
8730 8740 8750 8760 8770 8780 8790 8800
```

```
TACAGGCTTGGGATCAGGATCTGCTGGTCTATATACAGAATGGAATACAGATACCCAGTGGTTGATGATGATGAT
ThrAspCysGlyIleProAlaIleValProValValIleThrValAsnAspThrIleSerCysPheAspLeuGlyPhe
8810 8820 8830 8840 8850 8860 8870 8880
```

```
AAAGGCATTTACAAAGGAGCTAAGAGAAGAGAACACTGTTGAAATGAGATTGCTATCTTCCAGCTTTGACA
LysArgProLeuLysIleGluArgValThrLysTyrAlaProSerValProGluLeuArgValProArgLeu
8890 8900 8910 8920 8930 8940 8950 8960
```

```
AAGCGTTGGGATTAAATACAAAGGAGCTAAGAGAAGAGAACACTGTTGAAATGAGATTGCTATCTTCCAGCTTTGACA
ThrSerLeuIleAlaLysIleGlyIleGluIleGluIleThrSerThrThrValThrIleGluGluValAlaArg
8970 8980 8990 9000 9010 9020 9030 9040
```

```
AGCTAGGAGCGAGCTTCCAGGTGCTCCAGAAGGAAGATCTCAGATTTGCTGCTACGCATTGAGAAAAGCAAT
AlaLysAlaGluLeuArgLeuAspLeuHisGluGlyAspTyrProGluTrpLeuGluLeuGluGluAlaThr
```
Fig. 2. Complete nucleotide sequence of FeFV proviral DNA. The sequence starts at the 5' end of the 5'-LTR (position 1) and ends at the 3' end of the 3'-LTR (position 11657). The TATA box is shown in bold and the primer binding site (PBS), polypurine tract (PPT) and polyadenylation signal [poly(A)] are underlined.

Fig. 3. Hydrophilicity plots of the Gag protein from FeFV (A) and simian spumavirus (B). The plots were generated by MacVector version 5.0 (Kodak IBI) using the Kyte–Doolittle algorithm with a window size of 10 amino acids.
obtained using the other sequences were similar (data not shown), although there were fewer sequences available for comparison.

This work was supported by a project grant from the Wellcome Trust.

References


Renne, R., Friedl, E., Schweizer, M., Fleps, U., Turek, R. & Neumann-


Received 9 April 1997; Accepted 3 June 1997