JOURNAL OF VIROLOGY, Jan. 1990, p. 406–410
0022-538X/90/010406-05$02.00/0

Vol. 64, No. 1

Relationship of the *env* Genes and the Endonuclease Domain of the *pol* Genes of Simian Foamy Virus Type 1 and Human Foamy Virus

AYALEW MERGIA,* KAREN E. S. SHAW, JONATHAN E. LACKNER, AND PAUL A. LUCIW
Department of Pathology, School of Medicine, University of California, Davis, Davis, California 95616

Received 14 July 1989/Accepted 29 September 1989

We have molecularly cloned and sequenced a portion of the simian foamy virus type 1 (SFV-1); open reading frames representing the endonuclease domain of the polymerase (*pol*) and the envelope (*env*) genes were identified by comparison with the human foamy virus (HFV). Unlike the HFV genomic organization, the SFV-1 *pol* gene overlaps the *env* gene; thus, the open reading frames reported for HFV between *pol* and *env* is not present in SFV-1. Comparisons of predicted amino acid sequences of HFV and SFV-1 reveal that the endonuclease domains of the *pol* genes are about 84% related. The region predicted to encode the SFV-1 extracellular *env* domain is 569 codons; SFV-1 and HFV have 64% amino acid similarity in this *env* domain. The predicted hydrophobic transmembrane *env* proteins of both HFV and SFV-1 show about 73% similarity. A total of 16 potential glycosylation sites are found in SFV-1 *env*, and 15 are found in HFV; 11 are shared. SFV-1 has 25 cysteine residues, and HFV has 23 residues; all 23 cysteine residues of HFV are conserved in SFV-1. This sequence analysis reveals that the human and simian foamy viruses are highly related.

Spumavirinae (or foamy viruses), oncoviruses, and lentiviruses belong to three subfamilies of *Retroviridae*. Foamy viruses have been found in nonhuman primates (10, 11), cows (19), cats (5), hamsters (12), and humans (1); a clear connection of foamy viruses with disease has not been established. Several serologically distinct simian foamy viruses (SFVs) have been obtained from a variety of old-world (types 1 to 3) and new-world (types 4 and 8) monkeys, as well as from apes (types 6 and 7) and prosimians (type 5) (for a review, see reference 13). The relationship of these different serotypes of SFVs and the human foamy virus (HFV), whose genome was recently characterized (6, 21), remains to be determined. The genome of HFV is 12,085 bases in size and has four open reading frames (ORFs) in addition to the *gag*, *pol*, and *env* genes. The ORF designated SI is located at the intergenic region between *pol* and *env*. Three additional ORFs, *bel* 1, *bel* 2, and *bel* 3, are found 3′ to the *env* gene. Up to now, HFV is the only foamy virus whose genome has been cloned and sequenced. To characterize SFVs, we have molecularly cloned the genome of SFV type 1 (SFV-1) isolated from the macaque monkey and we have determined the nucleotide sequence of the *env* gene and the endonuclease domain of the *pol* gene.

The SFV-1 isolated from a rhesus (*Macaque mulatta*) monkey was kindly provided by Richard Heberling of the Southwest Foundation for Biomedical Research (San Antonio, Texas). Virus was propagated in the dog thymus cell line Cf2Th (23). High-molecular-weight DNA was prepared from Cf2Th cells at 8 days after infection, when cytopathic effects were noted. This DNA was digested with *Eco*RI restriction enzyme, electrophoresed on agarose gels, and blotted onto nitrocellulose membrane. A 45-base synthetic oligonucleotide representing a portion of the polymerase region of HFV (nucleotide sequence position 664 to 708 [6]), where the sequence has the most similarity with the *pol* genes of other retroviruses, was used as a probe (6). This probe detected an *Eco*RI restriction fragment at about 5.5 kilobases (kb). For molecular cloning, 50 μg of high-molecular-weight infected-cell DNA was digested to completion with *Eco*RI; it was electrophoresed on an agarose gel, and DNA from the size fraction corresponding to 4.5- to 6.5-kb fragments was recovered (20). This DNA was used to construct a library in the bacteriophage vector λgt10 (13), and the library was screened with the same HFV 45-base oligonucleotide probe that detected the 5.5-kb band on a Southern blot. The 5.5-kb SFV-1 DNA region was subsequently subcloned from the bacteriophage vector into the plasmid pUC18.

To demonstrate that the λgt10 recombinant that had been cloned from DNA of infected cells in fact contained sequences specific for SFV-1, whole-cell DNA isolated from uninfected and virus-infected dog thymus cells (Cf2Th) was probed with labeled cloned SFV-1 DNA under stringent annealing conditions. SFV-1-infected cell DNA gave positive annealing signals, whereas no signals was obtained with DNA from uninfected cells (our unpublished results). Viral-specific DNA fragments at 5.5 and 1.1 kb were noted. The recombinant 5.5-kb clone contains a portion of the 3′ long terminal repeat, and the SFV-1 band at 1.1 kb corresponds to a portion of the 5′ long terminal repeat of SFV-1 (our unpublished results). DNA sequences were determined by the dideoxy-chain termination method by using double-strand DNA template [α-^{35}S]dATP, and Sequenase polymerase enzyme (28) (U.S. Biochemicals, Cleveland, Ohio). pUC 18 with SFV-1 inserts was denatured and annealed with M13 sequencing primers. Additional primers were prepared by using the Pharmacia Gene Assembler (Pharmacia, Inc., Piscataway, N.J.) for automated oligonucleotide synthesis, and both strands of SFV-1 DNA were sequenced. The nucleotide sequence presented in Fig. 1 includes the endonuclease domain of the *pol* gene and the entire *env* gene of SFV-1. The predicted amino acid sequences were compared with the sequence of the HFV genome (6).

Computer analysis of the sequence upstream from the *pol* termination codon (position 582, Fig. 1) identified only a single ORF. Comparisons of predicted amino acid sequences of HFV and SFV-1 revealed that the endonuclease domains of the *pol* gene are about 84% related (Table 1). When conservative amino acid changes were included, the similarity increased to 92%. Sequence comparisons of the *pol* gene of HFV with lentiviruses and oncoviruses revealed no

* Corresponding author.

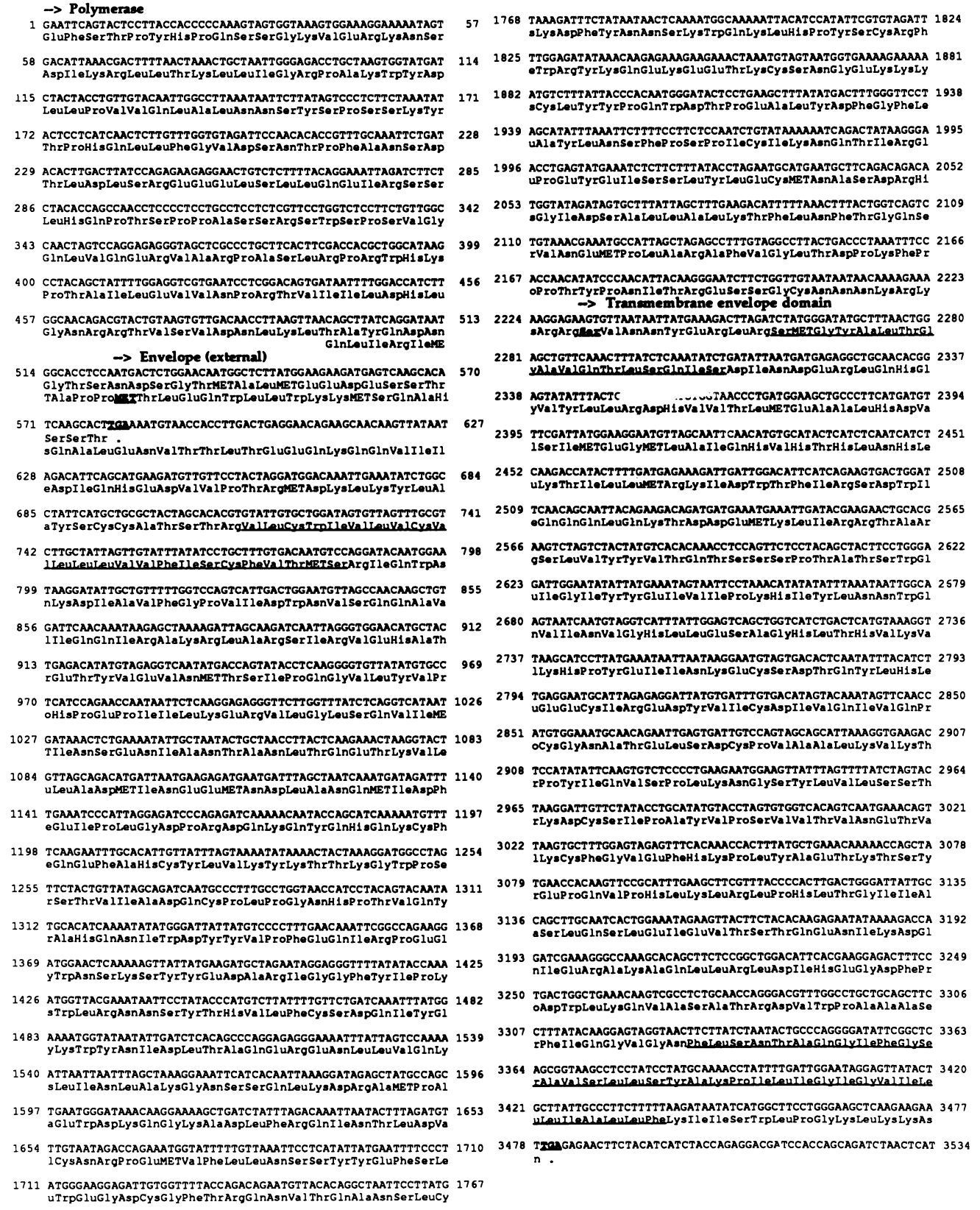

FIG. 1. DNA sequence of the endonuclease domain of the *pol* gene and the complete *env* gene of SFV-1. The DNA sequence has been numbered from the endonuclease domain of the *pol* gene to the end of the *env* gene. The predicted amino acid sequences for the *pol* gene and extracellular and transmembrane domains of the *env* gene are shown below the DNA sequence. Hydrophobic regions discussed in the text are underlined.

TABLE 1. Relatedness of human foamy virus and simian foamy virus type 1

Amino acid sequences	% Similarity[a]		
	pol gene (endonuclease)	External *env* domain	Transmembrane *env* domain
Conserved	84	64	73
Similar	92	80	88

[a] Comparisons, shown in percent, are based on predicted amino acid sequences.

significant similarity with members of either group (6, 21). The results presented here also strengthen the notion that foamy viruses belong to a separate subfamily of retroviridae. In the HFV genome, the intergenic region between the *pol* and *env* genes contains an ORF designated as S1, which could code for a polypeptide containing 107 amino acids. Although the SFV-1 nucleotide sequence is very similar to the HFV sequence in this region, we were unable to identify an ORF that corresponds to S1 of HFV after sequencing both strands of the SFV-1 DNA. Instead, the SFV-1 *pol* gene product overlaps the NH_2 terminus of the *env* gene by 19 amino acids. The *pol* and *env* genes of MuLV also overlap (30). It has not yet been established whether S1 as well as *bel* 1, *bel* 2, and *bel* 3 ORFs of HFV encode for any functional proteins. Introduction of site-specific mutations in these ORFs and analysis of specific viral messages by cDNA cloning are means to elucidate the significance of their products for virus replication. On the basis of the observation that the genomes of HFV and SFV-1 have similar sizes (18; our unpublished results), we anticipate SFV-1 to have extra ORFs that may be similar to the three *bel* ORFs of HFV.

An ORF encoding the *env* gene product of SFV-1 initiates in a different reading frame from *pol*, at a position 85 nucleotides (at position 498) upstream from the stop codon of *pol* (Fig. 1). The SFV-1 *env* gene contains 995 codons; a termination codon (TGA) is found at position 3534 (Fig. 1). A second ATG codon, with similarity to the consensus sequence for potential initiation codon (Met) (17), is located at residue 10 in the reading frame (nucleotide 525, Fig. 1). A second ATG, located 11 codons downstream in SFV-1, is also predicted to be the first codon for HFV *env*. Thus the *env* precursor polypeptide of SFV-1 is predicted to contain 986 amino acids. The putative initiation codon is followed by 62 amino acid residues that are hydrophilic (Fig. 1). The amino-terminal region of HFV, as well as visna (31), equine infectious anemia virus (15), and human immunodeficiency virus type 1 (22, 25, 27, 31), is also relatively hydrophilic. Like HFV, the first hydrophobic region of SFV-1 *env* comprises 24 amino acids (residues 64 to 87, Fig. 1). The first hydrophobic region has been proposed for visna (31), equine infectious anemia virus (15), and human immunodeficiency virus type 1 (22) to serve as a signal peptide for transport of the envelope glycoprotein precursor to the cytoplasmic membrane. The *env* gene precursor (gp160) of HIV-1 is processed by proteolytic cleavage within this region (2, 26).

A second hydrophobic region (17 residues at 2257 to 2308, Fig. 1) is preceded by the peptide Arg-Lys-Arg-Arg, which is similar to the site at which *env* gene products are processed (16, 27, 32) to give rise to the extracellular and transmembrane domains. The region predicted to encode the SFV-1 extracellular envelope domain is 569 codons, whereas HFV appears to specify a counterpart that is 568 codons. All of the 17 residues of the second hydrophobic region are conserved in both HFV and SFV-1. Amino acid comparison of the *env* gene products of HFV and SFV-1 shows that they are similar (Fig. 2, Table 1). The extracellular domains are 64% related, and, when conservative amino acid substitutions are taken into account, the homology is 80%. Although protein sequence data will be required to determine the precise N terminus of the mature extracellular domain, a potential amino-terminal signal could be identified by the hydrophilic stretch at the predicted initiation codon. In both SFV-1 and HFV, a potential proteolytic cleavage site, Arg-Lys-Arg-Arg, is found upstream from a hydrophobic region. A major surface *env* glycoprotein of SFV-1 appears to have a molecular weight of 70 kilodaltons (3). Processing from the cleavage site generates an unmodified extracellular domain of SFV-1 *env* product of about 63.5 kilodaltons (calculated to include the putative signal sequence), without accounting for carbohydrate residues. A total of 13 potential glycosylation sites (Asp-X-Ser/Thr) are found in the extracellular domain of SFV-1, and 12 are found in HFV; 8 of these are shared and 3 are located in close proximity (Fig. 2). All of the 16 cysteine residues in NH_2-terminal *env* domain of HFV are conserved in SFV-1; there are two additional cysteine residues in the latter (Fig. 2). The predicted extracellular domain of SFV-1 and HFV are 568 and 567 codons, respectively. Therefore, we predict that the major glycoproteins of SFV-1 and HFV are similar in size.

A third hydrophobic region, 36 amino acids long, probably corresponds to a putative transmembrane domain; a similar feature is noted in the *env* gene of HFV (Fig. 1). *env* genes of lentiviruses also have a third hydrophobic region in the transmembrane domain (15, 22, 25, 27, 31, 33) like HFV and SFV-1 have. The overall structure of the foamy virus *env* gene product appears to be strikingly similar to that of the lentiviruses. However, the external portion of transmembrane domain of SFV-1 and HFV is almost twice as large in other retrovirus subfamilies and may, therefore, fold in a unique structure that is distinct from that of other retroviruses. The predicted transmembrane domains of SFV-1 and HFV are 417 and 416 codons (6), respectively. Amino acid sequence comparisons of the transmembrane domains show a 73% similarity between the human and simian viruses (Fig. 2). The three potential glycosylation sites and the seven cysteine residues are conserved in the transmembrane domains of both viruses. HFV and SFV-1, thus, appear to be highly related.

Molecular characterization of foamy viruses is essential for elucidating several features of this subfamily of retroviruses, including regulation of viral gene expression, mechanism of latency, strain variation, and mechanism of cytopathology. Like the lentiviruses and certain oncogenic retroviruses (e.g. human T-cell leukemia virus type I), the primate foamy viruses contain ORFs in addition to the *gag*, *pol*, and *env* genes. These ORFs may specify proteins that regulate viral gene expression in a positive or negative fashion, like the transactivators for human immunodeficiency virus, simian immunodeficiency virus, and human T-cell leukemia virus type I (14, 24, 29). Interestingly, the relationship of SFV-1 and HFV with respect to degree of sequence divergence seems to be analogous to that of human immunodeficiency virus type 2 and simian immunodeficiency virus isolated from rhesus macaques (4, 7–9). Detailed molecular characterization of SFV isolates from several monkey species sharing a geographical area may provide insight on strain variation and interspecies spread of retroviruses. Thus, molecular investigations on foamy viruses can be directed at many issues that are relevant for

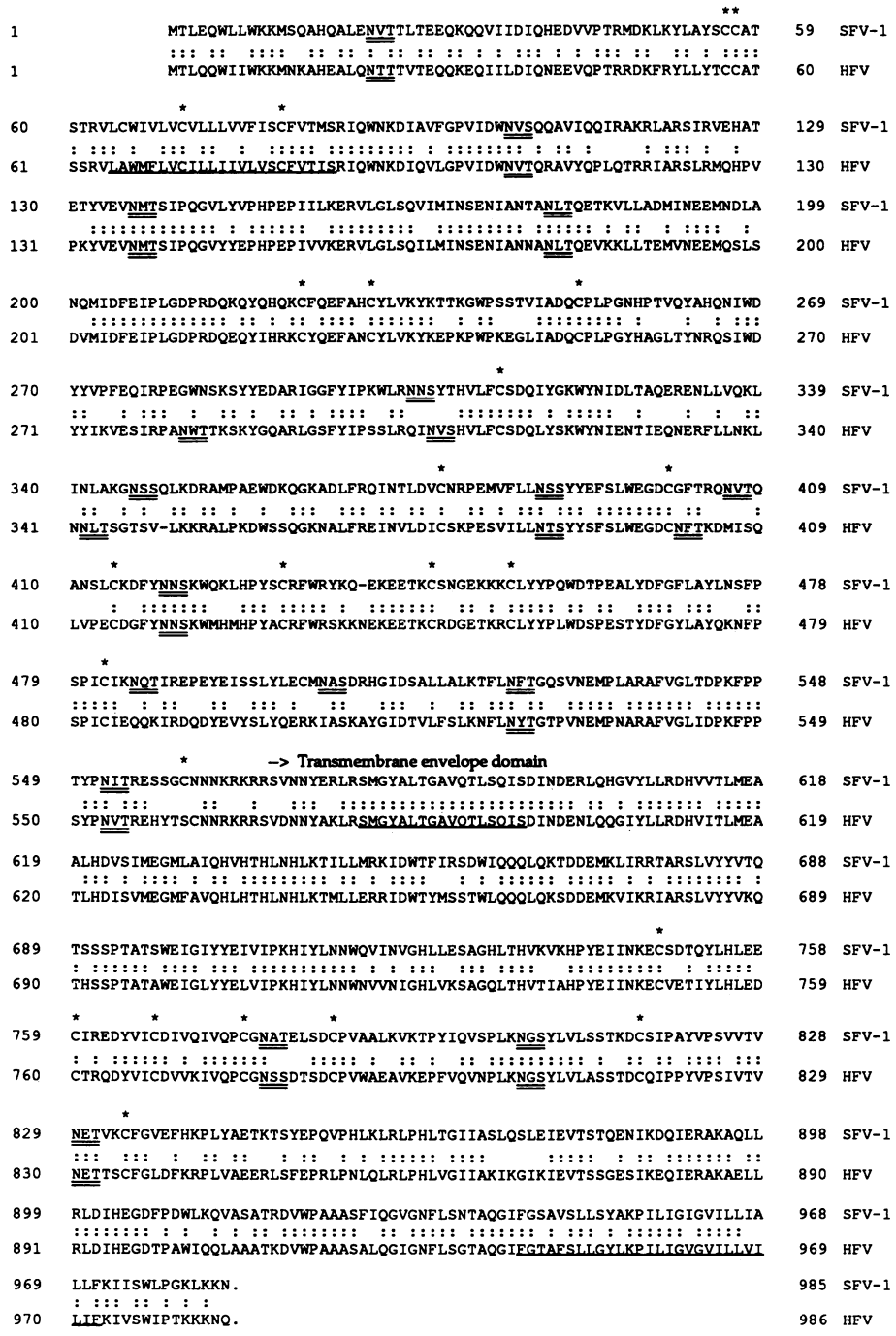

FIG. 2. Alignment of the predicted amino acid sequence encoded by the *env* genes of SFV-1 and HFV. Amino acids identical to both SFV-1 and HFV are indicated by two dots. The potential N-linked glycosylation sites of the type Asn-X-Ser/Thr are double underlined. The stars show shared cysteine residues between SFV-1 and HFV. Alignments were determined using the IFIND program provided by Intelligenetics Corp. (Mountain View, Calif.) to BIONET users. Hydrophobic domains discussed in the text are underlined.

understanding the interactions of susceptible hosts and pathogenic retroviruses.

We gratefully acknowledge Cyndi Russell and Warner Yuen for their technical assistance. We thank Preston Marx, Murray Gardner, and Niels Pedersen for helpful suggestions and advice.

This work was supported in part by a Center for AIDS Research grant (AI27732) to Murray Gardner from the National Institute of Allergy and Infectious Diseases. P.A.L. was a recipient of an Investigator Award from the California Universitywide Task Force on AIDS (K87D017).

LITERATURE CITED

1. **Achong, B. G., P. W. A. Mansell, M. A. Epstein, and P. Clifford.** 1971. Virus in cultures from a human nasopharyngeal carcinoma. J. Natl. Cancer Inst. **46**:299–307.
2. **Allan, J. S., J. E. Coligan, F. Barin, M. F. McLane, J. G. Sodroski, C. A. Rosen, W. A. Haseltine, T. H. Lee, and M. Essex.**

1985. Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. Science **228**:1091–1093.

3. **Benzair, A.-B., A. Rhodes-Feuillette, J. Lasneret, R. Emanoil-Ravier, and J. Peries.** 1985. Purification and characterization of the major envelope glycoprotein of simian foamy virus type 1. J. Gen. Virol. **66**:1449–1455.
4. **Chakrabarti, L., M. Guyader, M. Alizon, M. D. Daniel, R. C. Desrosiers, P. Tiollais, and P. Sonigo.** 1987. Sequence of simian immunodeficiency virus from macaque and its relationship to other human and simian retroviruses. Nature (London) **328**: 543–547.
5. **Fabricant, C. G., L. F. Rich, and J. H. Gillespie.** 1969. Feline viruses. XI. Isolation of a virus similar to a myxovirus from cats in which urolithiasis were experimentally induced. Cornell Vet. **59**:667–672.
6. **Flugel, R. M., A. Rethwilm, B. Maurer, and G. Darai.** 1987. Nucleotide sequence analysis of the *env* gene and its flanking regions of the human spumaretrovirus reveals two novel genes. EMBO J. **6**:2077–2084.
7. **Franchini, G., C. Gurgo, H.-G. Guo, R. C. Gallo, E. L. Collalti, K. A. Fargnoli, L. F. Hall, F. Wong-Staal, and M. S. Reitz, Jr.** 1987. Sequence of simian immunodeficiency virus and its relationship to the human immunodeficiency viruses. Nature (London) **328**:539–543.
8. **Guyader, M., M. Emerman, P. Sonigo, F. Clavel, L. Montagnier, and M. Alizon.** 1987. Genome organization and transactivation of the human immunodeficiency virus type 2. Nature (London) **326**:662–669.
9. **Hirsch, V. M., R. A. Olmsted, M. Murphey-Corb, R. H. Purcell, and P. R. Johnson.** 1989. An African primate lentivirus (SIVsm) closely related to HIV-2. Nature (London) **339**:389–392.
10. **Hooks, J. J., and B. Detrick-Hooks.** 1981. Spumavirinae: foamy virus group infections. Comparative aspects and diagnosis, p. 599–618. *In* E. Kurstak and C. Kurstak (ed.), Comparative diagnosis of viral disease, vol. 4. Academic Press, Inc., New York.
11. **Hooks, J. J., and C. G. Gibbs, Jr.** 1975. The foamy viruses. Bacteriol Rev. **39**:169–185.
12. **Hruska, J. F., and K. K. Takemoto.** 1975. Biochemical properties of a hamster syncytium (foamy) virus. J. Natl. Cancer Inst. **54**:601–605.
13. **Huynh, T. V., R. A. Young, and R. W. Davis.** 1985. Construction and screening cDNA libraries in lambda gt10 and lambda gt11, p. 49–78. *In* D. Glover (ed.), DNA cloning techniques: a practical approach, vol. 1. IRL Press Oxford.
14. **Inoue, J., M. Yoshida, and M. Seiki.** 1987. Transcriptional ($p40^{x}$) and post-transcriptional ($p27^{x\text{-}III}$) regulators are required for the expression and replication of human T-cell leukemia virus type I genes. Proc. Natl. Acad. Sci. USA **82**:3653–3657.
15. **Kawakami, T., L. Sherman, J. Dahlberg, A. Gazit, A. Yaniv, S. R. Tronick, and S. Aaronson.** 1987. Nucleotide sequence analysis of equine infectious anemia virus proviral DNA. Virology **158**:300–312.
16. **Kiyokawa, T., H. Yoshikura, S. Hattori, M. Secki, and M. Yoshida.** 1984. Envelope proteins of human T-cell leukemia virus: expression in *Escherichia coli* and its application to studies of *env* gene functions. Proc. Natl. Acad. Sci. USA **81**:6202–6206.
17. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
18. **Kupiec, J., J. Tapiero-Tabaly, M. Canivet, M. Santillana-Hayat, R. M. Flugel, J. Peries, and R. Emanoil-Ravier.** 1988. Evidence for a gapped linear duplex DNA in the replicative cycle of human and simian spumaviruses. Nucleic Acids Res. **16**:9557–9565.
19. **Malmquist, W. A., M. J. Van Der Maaten, and A. D. Boothe.** 1969. Isolation, immunodiffusion, immunofluorescence and electron microscopy of a syncytial virus of lymphosarcomatous and apparently normal cattle. Cancer Res. **16**:188–200.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Maurer, B., H. Bannert, G. Darai, and R. M. Flugel.** 1988. Analysis of the primary structure of the long terminal repeat and the *gag* and *pol* genes of the human spumaretrovirus. J. Virol. **62**:1590–1597.
22. **Muesing, M. A., D. H. Smith, C. D. Cabradilla, C. V. Benton, L. A. Lasky, and D. J. Capon.** 1985. Nucleic acid structure and expression of the human AIDS/lymphadenopathy retrovirus. Nature (London) **313**:450–458.
23. **Nelson-Rees, W. A., R. B. Owens, P. Arnstein, and A. J. Kniazeff.** 1976. Source, alterations, characteristics and use of a new dog cell line (Cf2Th). In Vitro **12**:665–669.
24. **Peterlin, B. M., and P. A. Luciw.** 1988. Molecular biology of HIV. AIDS Res. Hum. Retroviruses **2**:29–40.
25. **Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, R. Starcich, S. F. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Launfenberger, J. Papis, J. Ghrayeb, N. T. Chang, R. C. Gallo, and F. Wong-Staal.** 1985. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature (London) **313**:277–284.
26. **Robey, W. G., B. Safai, S. Oroszlan, L. O. Arthur, M. A. Gonda, R. C. Gallo, and P. J. Fischinger.** 1985. Characterization of envelope and core structural gene products of HTLV-III with sera, from AIDS patients. Science **228**:593–595.
27. **Sanchez-Pescador, R., M. D. Power, P. J. Barr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, A. Renard, A. Randolph, J. A. Levy, D. Dino, and P. A. Luciw.** 1985. Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). Science **227**:484–492.
28. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
29. **Seiki, M., S. Hattori, Y. Hirayama, and M. Yoshida.** 1983. Human adult T-cell leukemia virus: complete nucleotide sequence of the provirus genome integrated in leukemia cell DNA. Proc. Natl. Acad. Sci. USA **80**:3618–3622.
30. **Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe.** 1981. Nucleotide sequence of Moloney murine leukemia virus. Nature (London) **293**:543–548.
31. **Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Retzel, P. Tiollais, A. Hasse, and S. Wain-Hobson.** 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. Cell **42**:369–392.
32. **Veronese, F. D., A. L. DeVico, T. D. Copeland, S. Oroszlan, R. C. Gallo, and M. G. Sarngadharan.** 1985. Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. Science **229**:1402–1405.
33. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. Cell **40**:9–17.