The Visna Virus Genome: Evidence for a Hypervariable Site in the *env* Gene and Sequence Homology among Lentivirus Envelope Proteins

MICHAEL J. BRAUN,^{1†*} JANICE E. CLEMENTS,² AND MATTHEW A. GONDA¹

Laboratory of Cell and Molecular Structure, Program Resources, Inc., National Cancer Institute-Frederick Cancer Research Facility, Frederick, Maryland 21701,¹ and Departments of Neurology and Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, Maryland 21205²

Received 17 June 1987/Accepted 21 August 1987

The complete nucleotide sequence of the visna virus 1514 genome was determined. Our sequence confirms the relationship of visna virus and other lentiviruses to human immunodeficiency virus (HIV) both at the level of sequence homology and of genomic organization. Sequence homology is shown to extend to the transmembrane proteins of lentivirus *env* genes; this homology is strongest in the extracellular domain, suggesting that close structural and functional similarities may also exist among these envelope proteins. Comparison of our data with the sequence of visna virus LV1-1, an antigenic variant derived from strain 1514, demonstrates that the rate of divergence has been about 1.7×10^{-3} substitutions per nucleotide per year in vivo. This rate is orders of magnitude higher than that for most DNA genomes, but agrees well with estimates of the rate for HIV. A statistically significant cluster of mutations in the *env* gene appears to represent a hypervariable site and may correspond to the epitope responsible for the antigenic differences between 1514 and LV1-1. Analysis of the potential RNA folding pattern of the visna virus *env* gene shows that this hypervariable site falls within a region with little potential for intramolecular base pairing. This correlation of hypervariability with lack of RNA secondary structure is strengthened by the fact that it also holds for a hypervariable site in the *env* gene of HIV.

Visna virus, a retrovirus of sheep, is the best-known representative of the *Lentivirinae* subfamily of retroviruses (12). Lentiviruses are nononcogenic exogenous viruses which cause chronic, debilitating multisystem diseases with a slow onset in susceptible hosts. The discovery that the human immunodeficiency virus (HIV), the causative agent of human acquired immunodeficiency syndrome (AIDS), belongs to the *Lentivirinae* has broad implications for the AIDS disease process and has given new importance to the study of animal lentiviruses (2, 10, 11, 35, 38). In addition to visna virus and HIV, the lentivirus subfamily includes two other well-characterized retroviruses: caprine arthritis-encephalitis virus and equine infectious anemia virus (EIAV). Newly discovered or recently characterized isolates from cattle (M. A. Gonda, M. J. Braun, S. G. Carter, T. A. Kost, J. W. Bess, Jr., L. O. Arthur, and M. J. Van Der Maaten, *Nature* (London), in press), cats (28), and simian species (5) are likely candidates for inclusion in the lentivirus subfamily.

Lentiviruses preferentially infect cells of the immune system in vivo, although the specific target cell may differ. They undergo long periods of latency and then are expressed at high levels in subsets of cells which may be responding to cellular differentiation signals (24). Latency is probably the greatest factor in viral persistence (12), but antigenic variation (13, 22, 27) and the apparent inability of hosts to produce protective antisera with high titers (25, 40) may also contribute. Lentivirus infections can persist for months or years before the onset of symptoms. These diseases are

usually progressive and lead to death. Encephalitis as well as secondary infections are common. The encephalitis is presumably due to the neurotropism of the viruses. In vitro, these viruses are cytopathic, inducing syncytia which later lead to cell death. Most of these aspects of animal lentivirus biology have parallels in the AIDS disease process (10-12, 41).

To provide a framework for a comparative study of lentiviruses, we have determined the complete nucleotide sequence of the genome of the prototype visna virus 1514. Evidence presented in this report, and elsewhere (S. Gdovin and J. E. Clements, manuscript in preparation) supports the idea that the previously published visna virus sequence (35) is that of strain LV1-1 (23), an antigenic variant of strain 1514. Comparison of our 1514 sequence with that of LV1-1 reveals a cluster of mutations in the *env* gene which may allow LV1-1 to escape neutralization by antisera to 1514 (23). We also describe distant but recognizable sequence homology between the *env* genes of visna virus, HIV, and EIAV which may be of predictive value in attempts to disrupt lentivirus disease processes.

Sequence determination. The genome of visna virus 1514 was cloned as two *SacI* restriction fragments of 8.6 (21) and 0.5 kilobases (kb) (J. E. Clements, unpublished data). The 8.6-kb fragment was sequenced by the dideoxynucleotide chain termination method (33) from a random subclone library (7) in the M13 vector mp18 (26). The sequence of the intact fragment was reconstructed by using the computer programs of Staden (36). Each base was sequenced an average of 5.1 times, and 75% of the sequence was determined on both strands. The 0.5-kb *SacI* fragment was sequenced both by the chain termination and chemical degradation methods (20). The possibility of clustered *SacI* sites at the junction between the 0.5- and 8.6-kb fragments

* Corresponding author.

† Present address: Department of Biological Sciences, University of Cincinnati, Cincinnati, OH 45221.

TABLE 1. Neutralization of visna virus strains

Visna virus strains	Neutralizing antiserum titers		
	Anti-gp135 ^a	Sheep 1 ^b	Sheep 11 ^c
1514	1:320 ^d	1:320	1:160
LV1-1	<1:20	<1:20	1:160
Haase virus ^e	<1:20	<1:20	1:160

^a Antiserum made against purified gp135 from strain 1514.

^b Serum from sheep 1 inoculated with 1514. LV1-1 was isolated from this sheep at 2 years after infection. Early serum (before the appearance of LV1-1) was used for this neutralization test.

^c Serum was from a sheep inoculated with LV1-1.

^d Expressed as the highest dilution of serum which prevented virus-induced cytopathic effects in sheep choroid plexus cells (23). Each assay was performed twice, with the same result in every case.

^e Virus stock used in the derivation of the previously published visna virus DNA sequence (14, 35).

was eliminated by subcloning and sequencing, from an independent full-length visna virus proviral clone (derived by J. E. Clements; unpublished data), a 0.3-kb *HindIII-SmaI* fragment spanning this junction. The possibility of clustered *SacI* sites in the long terminal repeat (LTR) at the 5' end of the 0.5-kb fragment was ruled out by Sonigo et al. (35).

Comparison of visna virus sequences. A visna virus sequence has been previously published (35). This sequence, originally reported to represent visna virus 1514, was from a molecular clone derived independently in the laboratory of A. Haase (14) from visna virus stocks obtained from O. Narayan. Plaque-purified stocks of visna viruses 1514 and LV1-1 were immunologically authenticated in the laboratory of O. Narayan before being distributed to A. Haase in a single shipment. LV1-1 is one of several antigenic variants that arose in a sheep after experimental infection with visna virus 1514 (23). Gdovin and Clements (in preparation) have recently sequenced the *env* gene and 3' LTR of a clone of LV1-1 independently derived in their laboratory. Their LV1-1 sequence is identical to that reported by Sonigo et al. (35) for the Haase clone (14), whereas our sequence of strain 1514 differs by 12 nucleotide substitutions in the *env* gene and 3' LTR (see below). To clarify this paradox, we obtained from the Haase laboratory the visna virus stock used to derive the previously sequenced molecular clone and compared it with reference stocks of strains 1514 and LV1-1 in virus neutralization tests. Virus neutralization was performed as previously described (23). Equal volumes containing 100 50% tissue culture infective doses of virus and sequential dilutions of serum were mixed and incubated at 37°C for 30 min and then inoculated into four replicate cultures of sheep choroid plexus cells in microtiter plates. The Haase virus stock is antigenically identical to LV1-1 in these neutralization assays (Table 1). We therefore conclude that the previously published sequence (35) is actually that of LV1-1 and that the sequence reported herein is the first bona fide sequence of visna virus 1514, which is the progenitor of LV1-1 and other antigenic variants in wide use (23). All references in this paper to the sequence of LV1-1 pertain to the whole genome reported by Sonigo et al. (35) which Gdovin and Clements (in preparation) confirmed for the *env* gene and 3' LTR.

Our sequence of strain 1514 (Fig. 1) differs from that of LV1-1 (35) in 28 of 9,203 nucleotides (0.3% sequence divergence), as shown in Table 2. Our sequence confirms previous conclusions as to the architecture of the visna genome, including LTR structure and functional assignment of major open reading frames (35). Studies of ours and others on the relationship of HIV to the animal lentiviruses (2, 10, 11, 35,

38) are also confirmed. An important difference between the two visna virus molecular clones is that the *env* reading frame of strain 1514 is open throughout, whereas LV1-1 has a premature stop codon (Table 2). Nevertheless, attempts to demonstrate biological activity of the 1514 clone reported herein by transfection or microinjection of the cloned DNA into cells which support visna virus replication have not yet been successful, even though the cloned LTR activates transcription in vitro (15) and all reading frames of the major genes (*gag*, *pol*, *env*) are open.

We consider all 28 differences between the two visna virus clones to represent true sequence divergences, except possibly for the 1-base-pair (bp) deletion at position 104 (Table 2). This is the only insertion or deletion difference, and it occurs at an *EcoRII* site in our sequence. The sequence of *EcoRII* sites, CC(A or T)GG, often produces a pattern on chemical degradation sequencing gels in which one C is faint or missing. Since the previous sequence of this region (35) was determined by chemical degradation, we consider this difference to be a potential sequencing error.

Of the 27 nucleotide substitutions, transitions (purine to purine or pyrimidine to pyrimidine) outnumber transversions (purine to pyrimidine or vice versa) by nearly six to one (23 transitions, 4 transversions). A preponderance of transitions over transversions is often apparent in recently diverged sequences (8). Of the 27 substitutions, 26 occur in potential coding regions, and of these, 16 would result in amino acid replacements. Most of the predicted replacements are reasonably conservative; five however, result in charge changes (Table 2). Of the substitutions which occur in coding regions, five are in the first base of a codon, 12 are in the second base, and 10 are in the third base. This distribution does not differ significantly from the random expectation of an equal number of substitutions at each position (χ^2 test), even though nine of the third-base substitutions are synonymous and only one of the first- or second-base substitutions is synonymous. Evidently, a large enough percentage of the possible amino acid replacements is selectively equivalent such that the bias toward third-base substitutions usually caused by natural selection has yet to become apparent at this early stage of sequence divergence.

It is reasonable to assume that most of the sequence differences between our visna virus 1514 clone and the LV1-1 clone (35) arose during experimental infection of the host rather than during propagation of the virus stocks or subsequent to cloning. This idea is reinforced by the fact that no differences have arisen in the *env* gene or 3' LTR between the Haase LV1-1 clone (14) and that of Gdovin and Clements (in preparation), even though the two virus stocks were propagated independently for about 7 years before being cloned and amplified in bacteriophage λ . Thus, the 27 (or 28) mutations between 1514 and LV1-1 can be considered to have arisen in 1.75 years of evolutionary time (the period between infection with 1514 and isolation of LV1-1; 23). This implies an approximate average rate of sequence divergence of 1.7×10^{-3} substitutions per nucleotide per year for the entire visna virus genome during infection of a susceptible host. This value must be considered a rough approximation because of the short time span involved. Moreover, selection by the host immune system has undoubtedly favored some mutations. However, the estimated visna virus rate agrees well with values for the rate of HIV sequence divergence in vivo (10^{-4} for the *gag* gene and 10^{-3} for the *env* gene; 13). Compared with the rates for DNA genomes (19), however, these values are very high; they are only approached by the rates for some other RNA viruses (9, 16).

[illegible]

FIG. 1. The complete nucleotide sequence of visna virus 1514 in the form predicted for viral RNA, along with translation of the major open reading frames. Major structural features and restriction enzyme cleavage sites used for cloning are indicated and are discussed in the text. Potential glycosylation sites in the *env* gene are underlined. Where the nucleotide or predicted amino acid sequence differs from LVI-1, the LVI-1 residue (35) is shown above or below the line.

TABLE 2. Nucleotide sequence differences between visna viruses 1514 and LV1-1

Functional unit	Nucleotide position	Base in strain:		Amino acid in strain:	
		1514	LV1-1	1514	LV1-1
U5	104	C	None ^a	Noncoding	Noncoding
<i>gag</i>	551	T	C	Val	Ala
	838	A	G	Arg	Gly ^b
	1421	T	C	Met	Thr
	1629	A	C	Gln	His ^b
	1662	G	A	Gly	Gly
<i>pol</i>	2399	G	T	Pro	Pro
	2910	C	T	Leu	Leu
	3019	A	G	Lys	Arg
	3232	T	C	Ile	Thr
	3399	G	C	Val	Leu
	4430	A	G	Glu	Glu
	4679	C	T	Asn	Asn
<i>sor</i>	5076	G	A	Gln	Gln
<i>tat</i>	5701	C	T	Ala	Val
	5767	G	A	Gly	Asp ^b
<i>env</i>	5991	T	C	Met	Thr
	6309	A	G	Asn	Ser
	6804	G	A	Arg	Lys
	7459	C	T	Asn	Asn
	7756	C	T	Ser	Ser
	7864	A	G	Ser	Ser
	7875	C	A	Ala	Glu ^b
	7878	A	G	Gln	Arg ^b
	7890	A	G	Lys	Arg
	8074	A	G	Leu	Leu
	8162	C	T	Arg	Stop
U3	8966	G	A	Noncoding	Noncoding

^a Possibly a sequencing error. See text.^b This substitution results in a change in charge at this position.

Therefore, the overall potential for rapid evolution of lentiviruses is great.

Hypervariable site in the *env* gene. The distribution of substitutions throughout the genome appears random except for one cluster of mutations in the *env* gene (Fig. 2). This cluster occurs near the putative COOH terminus of the outer membrane protein (OMP) and includes one synonymous and three nonsynonymous mutations in a 27-bp stretch (nucleotides 7864 through 7890; Fig. 1). To estimate the probability of this cluster occurring by chance, a series of Monte Carlo simulations were performed. In each simulation, the genome was divided into X segments of length (9,203/X bp). Then a large number of iterations was performed, with each iteration consisting of the computer generation of 28 random numbers between 1 and X to represent the 28 observed mutations. The frequency of iterations in which 4 or more of the 28 numbers were the same was taken as an estimate of the probability of four or more mutations occurring by chance in one of the X genome segments. It was initially felt that choosing genome segments of exactly 27 bp might bias the results, so several simulations were run. Three simulations of 10,000 iterations each were run with 100 genome segments of 92.03 bp. Each simulation used a different random number generator. These simulations converged on a value of 0.012 (range 0.0120 to 0.0123) as the probability of having four mutations in a 92-bp stretch, given 28 mutations

overall. Then a simulation of 150,000 iterations was run with 300 genome segments of 30.67 bp each, using the GGUD random number generator from the International Mathematical and Statistical Library. This simulation produced a value of 0.000513, or about 5 chances in 10,000 of having four mutations in a 30-bp stretch, given 28 mutations overall. We therefore concluded it unlikely that all the mutations observed in the visna virus genome were distributed randomly and that the cluster of mutations in the OMP of the *env* gene may be due to a tendency for mutations to accumulate in this region.

Genetic variation during the course of infection of an individual is a remarkable feature of lentivirus diseases. Such variation has been observed for visna virus, EIAV, and HIV (13, 22, 27) and has been shown to correlate with changes in antigenicity in the case of visna virus and EIAV. It has consistently been observed that the preponderance of mutations in natural lentivirus variants occurs in the *env* gene, especially in the portion coding for the OMP (3, 13, 31). A logical explanation for this distribution of mutations is that selection by the host immune system against the infecting antigenic strain of virus confers a temporary growth advantage on these newly arisen *env* variants, which initially escape the immune response. However, this immune selection mechanism predicts only a generalized variability throughout the *env* gene relative to other viral genes.

The very tight clustering of mutations seen in the visna virus *env* gene, in which 15% of all LV1-1 substitutions occur in a contiguous stretch representing only 0.3% of the genome, invites speculation on two other possibilities. First, it seems likely that this cluster of mutations marks the epitope responsible for the antigenic differences between 1514 and LV1-1, because there are only three other nonsynonymous substitutions in *env* (Table 2). Second, it seems plausible that this hypervariable site may also be a hypermutable site in the sense that more mutations actually occur at this site, whether or not the mutations later become prevalent because they are selectively advantageous to the virus. Cloning and sequencing of additional visna virus variants may provide information in this respect.

Secondary structure of visna virus genomic RNA. The observation of an apparently hypermutable site in the visna virus *env* gene led us to consider physicochemical mechanisms which might cause this mutability. We considered the possibility that the hypervariable site lay in a region of unusual secondary structure in the visna virus genomic RNA, which might in some way increase the rate of mutation. We examined this possibility by using the FOLD program (42) to predict the most stable secondary structure of a 500-bp segment of RNA surrounding the hypervariable site (bases 7675 to 8174). This segment was centered on the putative processing site between the OMP and transmembrane protein (TMP) and includes 250 bp on either side. On

FIG. 2. Distribution of nucleotide substitutions between visna viruses 1514 and LV1-1 (35). The positions of the major open reading frames are shown (—). The arrow indicates the OMP-TMP cleavage site in the *env* gene. Each square represents a single nucleotide substitution.

FIG. 3. Potential secondary structure of a 500-nucleotide RNA segment extending 250 nucleotides on either side of the OMP-TMP cleavage site in the *env* genes of visna virus 1514 and four strains of HIV (29, 32, 37, 39). Secondary structure predictions were made by using the FOLD program (42). Arrowheads indicate the position of a hypervariable site in each segment. The hypervariable site for HIV strains lies at nucleotides 7170 through 7196 of the BH10 isolate (29) and was identified by Hahn et al. (13); that for visna virus lies at nucleotides 7864 through 7890 (Fig. 1).

the basis of this analysis, the hypervariable site falls in a large, looped-out structure predicted to have little or no intramolecular base pairing (Fig. 3).

For comparison, we then predicted secondary structures for the same segments (250 bp on either side of the OMP-TMP processing site) of four HIV isolates (Fig. 3). Several hypervariable sites in the HIV *env* gene have been reported (13, 37); one of these lies in the segment we studied at nucleotides 7170 through 7196 of the BH10 isolate (29). In three of four cases, the hypervariable site either fell within or overlapped a major looped-out region of the predicted secondary structure and the fourth overlapped a minor looped-out region (Fig. 3).

The correlation among hypervariable sites and regions predicted to have little intramolecular base pairing in both the visna virus and HIV genomes may provide a clue to mechanisms inducing mutability. The apparent lack of base pairing may indicate that there are no secondary structural constraints in this region which might limit the rate of sequence divergence. Alternatively, this region might be more exposed to chemical mutagens than regions with more secondary structure. Moreover, single-stranded regions may appear more mutable because mutations that occur there are not deleterious to survival. Finally, it may be more difficult for reverse transcriptase (or other polymerases) to replicate these regions accurately, although the lack of secondary structure might lead one to expect the opposite. In this respect, it is possible therefore that the difficulty in replicating the unstructured areas may be that they form pseudoknots with other regions of the genome which were not included in the analysis. Whatever the physical basis, we entertain the notion that a lack of intramolecular base pairing at the RNA level may contribute in some way to the hypervariability observed at these sites in HIV and visna virus. Although the idea is speculative, it is noteworthy that higher rates of substitution in loops also have been observed in the poliovirus genome, another RNA virus (4).

Regardless of the exact mechanism, the advantage to the virus of having a hypermutable *env* gene is clear. The *env* glycoproteins, especially the OMP, are the antigens against which neutralizing antibodies are believed to be directed. Any viral variant in which the *env* gene mutates frequently would have a growth advantage, because it could escape elimination by the host immune response to the original viral serotype. Variants in which replication was error prone throughout the genome could achieve frequent *env* mutations, but only at the cost of introducing many, often deleterious, mutations in other genes. A more efficient mechanism would be one in which only the *env* gene and, perhaps, only antigenic sites within the *env* gene, are hypermutable. Viral variants arising in vitro in the absence of host immune response will have to be examined for direct evidence that the latter sort of mechanism may be operative.

Homologies among lentivirus *env* proteins. Previous studies have elucidated the similarities among the *gag* and *pol* gene products of various lentiviruses, but no sequence homologies between *env* proteins have been described. We compared the translated protein sequences of visna virus, HIV, and EIAV *env* genes by using the ALIGN computer program (6). This program uses the Needleman-Wunsch algorithm and an empirically derived mutation data scoring matrix to determine the maximum match score possible for two sequences and an alignment with that score. The program then estimates the significance of the alignment by comparing the actual match score with the average match score of a number of random permutations of the two sequences. An alignment score is computed which represents the number of standard deviations (SD) above the random average that an actual match score lies.

When the OMP-TMP processing site is used as a benchmark to align the *env* genes of visna virus, HIV, and EIAV, significant homology can be seen (Fig. 4). This homology extends 5' for about 60 amino acids into the OMP before falling off, but the homology is more extensive in the TMP.

FIG. 4. Alignment of the extracellular and transmembrane domains of the transmembrane *env* proteins of visna virus 1514, HIV (29), and EIAV (30). The alignment was performed by using the ALIGN program (6), with a gap penalty of 25, a matrix bias of 0, and 300 random comparisons. Identical residues are indicated (●). The alignments shown are those optimal with HIV; the visna virus-EIAV alignment can be improved slightly by moving gaps. Potential N-linked glycosylation sites are boxed. Arrows indicate conserved cysteine residues.

FIG. 5. Hydrophobicity profiles of two lentivirus *env* TMPs (visna virus and HIV; 29), a murine type C TMP (Moloney murine leukemia virus [Mo-MuLV]; 34), and a type D TMP (baboon endogenous virus [BAEV]). Baboon endogenous virus is a type C virus particle on the basis of morphology, but its *env* gene shows strong sequence homology (80% identity in TMP) with type D viruses (Cohen and Braun, unpublished data); therefore, it has been used here to represent type D TMPs. Hydrophobic regions appear above the mean; hydrophilic regions appear below the mean. The analysis was performed by the method of Hopp and Woods (17), with a step size of 1 and an averaging window of 7 for the lentivirus TMPs and 10 for the types C and D TMPs. The transmembrane domains, predicted by the method of Klein et al. (18), are indicated. The hydrophobicity profile of the EIAV TMP (not shown) also has features characteristic of lentivirus TMPs, as discussed in the text.

Retroviral TMPs generally contain three functional domains: an amino-terminal extracellular domain, a hydrophobic transmembrane domain, and a hydrophilic intracellular anchor. Sequence identity in the TMP among visna virus, HIV, and EIAV is about 20% in the extracellular domain and reaches 43% in the transmembrane domain of the visna virus-HIV comparison. Furthermore, many nonidentical amino acids are chemically similar, and these also contribute to the alignment score. We consider this homology good, especially in light of the fact that some HIV isolates differ in their TMPs by a 15% mismatch (1). A rough approximation of the statistical significance of the homology in the extracellular domain can be estimated from the alignment scores of 11.13 SD for visna virus versus HIV, 7.64 SD for HIV versus EIAV, and 6.74 SD for visna virus versus EIAV. An alignment score of greater than 6 SD is considered highly significant (6). Furthermore, this alignment preserves the position of a number of potentially important structural features in the proteins, including potential N-linked glycosylation sites and two cysteine residues (the only two cysteines in the extracellular domain of HIV). All of these features are well conserved in a number of HIV isolates (1). In the intracellular region 3' of the transmembrane domain, however, the sequence homology falls to insignificant levels, with alignment scores ranging from 0.4 to -2.4 SD.

In addition to overall alignment and conservation of structural features, the hydrophobicity profiles of these TMPs are very similar (Fig. 5). They all have major amino-terminal and transmembrane hydrophobic domains that are similarly positioned in the protein sequences. There is also a close correlation among the alternating hydrophilic and hydrophobic segments in the extracellular domains (residues 53 through 174). This resemblance is in contrast to hydrophobicity profiles of the TMPs of types C or D viruses, which resemble each other but not the lentiviruses (Fig. 5). Interestingly, in the cytoplasmic domains (residues 200 to the end), in which the amino acid homology is poor, there is little similarity in the hydrophobicity profiles. Nevertheless, the cytoplasmic domains of visna virus, HIV, and EIAV still share their overall hydrophilic character and a size larger than that for any other sequenced retrovirus (130 to 220

amino acids). To the foregoing similarities may be added the fact that the visna virus, HIV, and EIAV *env* proteins are both larger and more highly glycosylated than those of any other retroviral *env* proteins (30). Taken together, the sequence homology and structural similarities among lentivirus *env* genes provide strong evidence that these genes are ancestrally related and that functional similarities may be expected. Hence, in addition to HIV, other lentivirus *env* genes may provide good models for testing interventive strategies in the prevention of HIV infection and AIDS.

We thank J. Elser, J. Shumaker, P. Dorn-Williams, and G. LeBlonde for skilled technical assistance; J. Owens, G. Alvord, R. Stephens, and D. Lomb for help with computer and statistical analyses; and J. Hopkins for preparing the manuscript.

The research was sponsored, at least in part, by a Public Health Service grant from the National Cancer Institute under contract NO1-CO-23910 with Program Resources, Incorporated (to M.J.B. and M.A.G.) and by Public Health Service grants NS16145 and NS21916 from the National Institutes of Health (to J.E.C.).

LITERATURE CITED

- Alizon, M., S. Wain-Hobson, L. Montagnier, and P. Sonigo. 1986. Genetic variability of the AIDS virus: nucleotide sequence analysis of two isolates from African patients. *Cell* 46:63-74.
- Chiu, I.-M., A. Yaniv, J. E. Dahlberg, A. Gazit, S. F. Skuntz, S. R. Tronick, and S. A. Aaronson. 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature (London)* 317:366-368.
- Clements, J. E., N. D'Antonio, and O. Narayan. 1982. Genomic changes associated with antigenic variation of visna virus. *J. Mol. Biol.* 158:415-434.
- Currey, K. M., B. M. Peterlin, and J. V. Maizel, Jr. 1986. Secondary structure of poliovirus RNA: correlation of computer-predicted with electron microscopically observed structure. *Virology* 148:33-46.
- Daniel, M. D., N. L. Letvin, N. W. King, M. Kannagi, P. K. Sehgal, R. D. Hunt, P. J. Kanki, and R. C. Desrosiers. 1985. Isolation of T-cell lymphotropic HTLV-III-like retrovirus from macaques. *Science* 228:1201-1204.
- Dayhoff, M. O., W. C. Barker, and L. T. Hunt. 1983. Establishing homologies in protein sequences. *Methods Enzymol.* 91:524-545.
- Deininger, P. L. 1983. Random subcloning of sonicated DNA: application to shotgun DNA sequence analysis. *Anal. Biochem.* 129:216-223.
- Gojobori, T., W.-H. Li, and D. Graur. 1982. Patterns of nucleotide substitutions in pseudogenes and functional genes. *J. Mol. Evol.* 18:360-369.
- Gojobori, T., and S. Yokoyama. 1985. Rates of evolution of the retroviral oncogene of Moloney murine sarcoma virus and of its cellular homologues. *Proc. Natl. Acad. Sci. USA* 82:4198-4201.
- Gonda, M. A., M. J. Braun, J. E. Clements, J. M. Pyper, R. C. Gallo, F. Wong-Staal, and R. V. Gilden. 1986. HTLV-III shares sequence homology with a family of pathogenic lentiviruses. *Proc. Natl. Acad. Sci. USA* 83:4007-4011.
- Gonda, M. A., F. Wong-Staal, R. C. Gallo, J. E. Clements, O. Narayan, and R. V. Gilden. 1985. Sequence homology and morphologic similarity of HTLV-III and visna virus, a pathogenic lentivirus. *Science* 227:173-177.
- Haase, A. T. 1986. Pathogenesis of lentivirus infections. *Nature (London)* 322:130-136.
- Hahn, B. H., G. M. Shaw, M. E. Taylor, R. R. Redfield, P. D. Markham, S. Z. Salahuddin, F. Wong-Staal, and R. C. Gallo. 1986. Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science* 232:1548-1553.
- Harris, J. D., H. Blum, J. Scott, B. Traynor, P. Ventura, and A. Haase. 1984. The slow virus visna: reproduction *in vitro* of virus from extrachromosomal DNA. *Proc. Natl. Acad. Sci. USA* 81:7212-7215.
- Hess, J. L., J. E. Clements, and O. Narayan. 1985. *cis*- and *trans*-acting transcriptional regulation of visna virus. *Science* 229:482-485.
- Holland, J., K. Spindler, F. Horodyski, E. Grabau, S. Nichol, and S. VandePol. 1982. Rapid evolution of RNA genomes. *Science* 215:1577-1585.
- 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.
- Klein, P., M. Kanehisa, and C. DeLisi. 1985. The detection and classification of membrane-spanning proteins. *Biochim. Biophys. Acta* 815:468-476.
- Li, W.-H., T. Gojobori, and M. Nei. 1981. Pseudogenes as a paradigm of neutral evolution. *Nature (London)* 292:237-239.
- Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* 65:499-560.
- Molineaux, S., and J. E. Clements. 1983. Molecular cloning of unintegrated visna viral DNA and characterization of frequent deletions in the 3' terminus. *Gene* 23:137-148.
- Narayan, O., D. E. Griffin, and J. Chase. 1977. Antigenic drift of visna virus in persistently infected sheep. *Science* 197:376-378.
- Narayan, O., D. E. Griffin, and J. E. Clements. 1978. Virus mutation during "slow infection": temporal development and characterization of mutants of visna virus recovered from sheep. *J. Gen. Virol.* 41:343-352.
- Narayan, O., S. Kennedy-Stoskopf, D. Sheffer, D. E. Griffin, and J. E. Clements. 1983. Activation of caprine arthritis-encephalitis virus expression during maturation of monocytes to macrophages. *Infect. Immun.* 41:67-73.
- Narayan, O., D. Sheffer, D. E. Griffin, J. Clements, and J. Hess. 1984. Lack of neutralizing antibodies to caprine arthritis-encephalitis lentivirus in persistently infected goats can be overcome by immunization with inactivated *Mycobacterium tuberculosis*. *J. Virol.* 49:349-355.
- Norrand, J., T. Kempe, and J. Messing. 1983. Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene* 26:101-106.
- Payne, S., B. Parekh, R. C. Montelaro, and C. J. Issel. 1984. Genomic alterations associated with persistent infections by equine infectious anaemia virus, a retrovirus. *J. Gen. Virol.* 65:1395-1399.
- Pedersen, N. C., E. W. Ho, M. L. Brown, and J. K. Yamamoto. 1987. Isolation of a T-lymphotropic virus from cats with an immunodeficiency-like syndrome. *Science* 235:790-793.
- Ratner, L., W. Haseltine, R. Patarca, K. J. Livak, B. Starcich, S. J. Josephs, E. R. Doran, J. A. Rafalski, E. A. Whitehorn, K. Baumeister, L. Ivanoff, S. R. Petteway, Jr., M. L. Pearson, J. A. Lautenberger, T. S. Papas, J. Ghayab, 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.
- Rushlow, K., K. Olsen, G. Stiegler, S. L. Payne, R. C. Montelaro, and C. J. Issel. 1986. Lentivirus genomic organization: the complete nucleotide sequence of the *env* gene region of equine infectious anemia virus. *Virology* 155:309-321.
- Salinovich, O., S. L. Payne, R. C. Montelaro, K. A. Hussain, C. J. Issel, and K. L. Schnorr. 1986. Rapid emergence of novel antigenic and genetic variants of equine infectious anemia virus during persistent infection. *J. Virol.* 57:71-80.
- Sanchez-Pescador, R., M. D. Power, P. J. Barr, K. S. Steimer, M. M. Stempien, S. L. Brown-Shimer, W. W. Gee, and P. A. Luciw. 1985. Nucleotide sequence and expression of the AIDS-associated retrovirus (ARV-2). *Science* 227:484-492.
- Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74:5463-5468.
- Shinnick, T. M., R. A. Lerner, and J. G. Sutcliffe. 1981. Nucleotide sequence of Moloney murine leukaemia virus. *Nature (London)* 293:543-548.
- Sonigo, P., M. Alizon, K. Staskus, D. Klatzmann, S. Cole, O. Danos, E. Rétzel, and S. Wain-Hobson. 1985. Nucleotide sequence of the visna lentivirus: relationship to the AIDS virus. *Cell* 42:369-382.
- Staden, R. 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing.

- ing. *Nucleic Acids Res.* **10**:4731-4751.
37. **Starcich, B. R., B. H. Hahn, G. M. Shaw, P. D. McNeely, S. Modrow, H. Wolf, E. S. Parks, and W. P. Parks.** 1986. Identification and characterization of conserved and variable regions in the envelope gene of HTLV-III/LAV, the retrovirus of AIDS. *Cell* **45**:637-648.
38. **Stephens, R. M., J. W. Casey, and N. R. Rice.** 1986. Equine infectious anemia virus *gag* and *pol* genes: relatedness to visna and AIDS virus. *Science* **231**:589-594.
39. **Wain-Hobson, S., P. Sonigo, O. Danos, S. Cole, and M. Alizon.** 1985. Nucleotide sequence of the AIDS virus, LAV. *Cell* **40**:9-17.
40. **Weiss, R. A., P. R. Clapham, R. Cheingsong-Popov, A. G. Dalgleish, C. A. Carne, I. V. D. Weller, and R. S. Tedder.** 1985. Neutralization of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients. *Nature (London)* **316**:69-72.
41. **Wong-Staal, F., and R. C. Gallo.** 1985. Human T-lymphotropic retroviruses. *Nature (London)* **317**:395-403.
42. **Zuker, M., and P. Stiegler.** 1981. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res.* **9**:133-148.