

Identification and Characterization of Conserved and Variable Regions in the Envelope Gene of HTLV-III/LAV, the Retrovirus of AIDS

Bruno R. Starcich,* Beatrice H. Hahn,†
George M. Shaw,† Paul D. McNeely,†
Susanne Modrow,‡ Hans Wolf,‡ Elizabeth S. Parks,§
Wade P. Parks,§ Steven F. Josephs,* Robert C. Gallo,*
and Flossie Wong-Staal*

*Laboratory of Tumor Cell Biology
Division of Cancer Treatment
National Cancer Institute
Bethesda, Maryland 20205

†Division of Hematology and Oncology
University of Alabama Medical Center
Birmingham, Alabama 35294

‡Molecular and Tumor Virology
Max von Pettenkofer Institute
D 8000 Munich 2

West Germany

§Department of Pediatrics
University of Miami School of Medicine
Miami, Florida 33101

Summary

To determine the extent and nature of genetic variation present in independent isolates of HTLV-III/LAV, the nucleotide sequences of the entire envelope gene and parts of *gag* and *pol* were determined for two AIDS viruses. The results indicated that variation throughout the viral genome is extensive and that the envelope gene in particular is most highly variable. Within the envelope, changes were most prevalent within the extracellular region where clustered nucleotide substitutions and deletions/insertions were evident. Based on predicted secondary protein structure and hydrophilicity, these hypervariable regions represent potential antigenic sites. In contrast to the hypervariable regions, other sequences in the extracellular envelope and the overall envelope structure (including 18 of 18 cysteine residues), as well as most of the transmembrane region, were highly conserved.

Introduction

The extreme morbidity, high mortality, and epidemic proportions of the acquired immune deficiency syndrome (AIDS) have led to intense scientific efforts to elucidate the disease's pathogenesis and develop effective preventive and treatment measures. Fundamental to this effort was the identification and characterization of HTLV-III/LAV as the causative agent of AIDS (Barre-Sinoussi et al., 1983; Popovic et al., 1984; Gallo et al., 1984; Schupbach et al., 1984; Sarngadharan et al., 1984) and the subsequent molecular characterization of the viral genome (Hahn et al., 1984; Shaw et al., 1984; Alizon et al., 1984; Luciw et al., 1984; Ratner et al., 1985a; Wain-Hobson et al., 1985; Sanchez-Pescador et al., 1985; Muesing et al., 1985).

A troublesome finding, given its potential clinical impli-

cations, made early in the genetic analysis of HTLV-III/LAV was that independent virus isolates were heterogeneous in their genomic restriction enzyme cleavage patterns (Shaw et al., 1984, 1985; Wong-Staal et al., 1985). This observation was extended by electron microscopic heteroduplex studies (Hahn et al., 1985), DNA sequencing (Ratner et al., 1985a; Wain-Hobson et al., 1985; Sanchez-Pescador et al., 1985; Muesing et al., 1985), and additional restriction enzyme mapping (Bønn et al., 1985), all of which indicated that substantial genomic variation is present in different AIDS virus isolates and that the envelope might be particularly variable compared to other viral genes (Ratner et al., 1985b; Rabson and Martin, 1985). From these studies, however, it was impossible to estimate accurately the extent and nature of genomic heterogeneity present among independent AIDS virus isolates because restriction mapping and heteroduplexing lacked sufficient sensitivity and because sequence comparisons were performed on only three virus isolates, two of which were relatively closely related to each other.

Because of these limitations, and because a more complete understanding of genomic heterogeneity would undoubtedly be critical to elucidating the virus's evolutionary history and pathobiology, we undertook a study of genetic variation in five independent HTLV-III/LAV isolates. Here we present an analysis of the nucleotide and deduced amino acid sequences of the complete envelope genes and parts of *gag* and *pol* for HTLV-III/LAV isolates obtained from two Haitian patients with AIDS and compare these to the published sequences of HTLV-III (Ratner et al., 1985a), LAV (Wain-Hobson et al., 1985), and ARV (Sanchez-Pescador et al., 1985). The results of these studies provide considerable new insight into the extent and nature of genetic variation present in field isolates of HTLV-III/LAV, and provide important information regarding the rate and mechanisms by which these changes occur in vivo.

Results

Nucleotide Sequence Determination and Analysis

Isolates of HTLV-III/LAV were obtained from peripheral blood mononuclear cells. Isolate RF was obtained in 1983 as previously described (Popovic et al., 1984) and was propagated in an immortalized T cell line from which it was molecularly cloned (Hahn et al., 1985). Isolate WMJ-1 was similarly transmitted from the peripheral blood mononuclear cells of a Haitian infant with AIDS to an immortalized T cell line (Hahn et al., 1986). Molecular clones of both RF (designated HAT-3) and WMJ-1 were prepared as described in Experimental Procedures.

The genome of HTLV-III/LAV consists of long terminal repeat (LTR) elements; *gag*, *pol*, and *env* genes; two regions of unknown significance designated *so* for short open reading frame and *3' orf* for 3' open reading frame; and a third region previously thought to be noncoding (NCR) but now known to encode the first exon of the HTLV-

Table 1. Sequence Comparison of Five Independent AIDS Virus Isolates

Nucleotide Differences Compared to BH-10 (number of nucleotide changes/number of nucleotides sequenced)												
Clones	LTR	Leader Sequence and Primer Binding Site						envelope				Total
			<i>gag</i>	<i>pol</i>	<i>sor</i>	NCR (<i>tat</i>)	Signal Peptide	Extra-cellular portion	Trans-membrane portion	3' orf		
HAT-3	45/634 7%	13/152 8.5%	77/1285 5.9%	46/1246 3.7%	17/464 3.6%	48/584 8.2%	26/111 23.4%	202/1443 14.0%	87/1035 8.4%	69/648 10.6%	645/7489 8.6%	
ARV-2	30/634 4.7%	14/152 9.2%	86/1536 5.6%	134/3045 4.4%	31/609 5.1%	49/584 8.4%	26/111 23.4%	164/1443 11.4%	66/1035 6.3%	52/648 8.0%	582/9213 6.3%	
LAV-1a	10/634 1.6%	7/152 4.6%	46/1536 3.0%	59/3045 1.9%	2/609 0.3%	11/584 1.9%	2/111 1.8%	32/1443 2.2%	9/1035 0.9%	13/648 2.0%	144/9213 1.5%	
WMJ-1	nd	nd	38/1162 ^a 3.3%	nd	nd	nd	27/111 24.3%	177/1443 12.3%	62/1035 6.0%	nd	nd	
Amino Acid Differences Compared to BH-10 (number of amino acid changes/number of amino acids sequenced)												
HAT-3	—	—	28/428 6.5%	22/415 5.3%	10/154 6.4%	—	11/37 29.7%	103/481 21.4%	45/345 13.0%	37/216 17%	256/2039 12.5%	
ARV-2	—	—	32/512 6.3%	51/1015 5.0%	20/203 9.8%	—	13/37 35.1%	82/481 17.0%	42/345 12.2%	29/216 12.2%	239/2593 9.2%	
LAV-1a	—	—	16/512 3.1%	21/1015 2.1%	0/203 0%	—	2/37 5.4%	14/481 2.9%	5/345 1.4%	8/216 3.7%	58/2593 2.2%	
WMJ-1	—	—	12/387 ^a 3.1%	nd	nd	—	13/37 35.1%	90/481 18.7%	29/345 8.4%	nd	nd	

Nucleotide and amino acid sequence comparisons of five independent AIDS virus isolates using the BH-10 clone of HTLV-IIIb as a reference (Ratner et al., 1985a). Deletions and insertions, as illustrated in Figures 1 and 2, are included in the analysis. Data for ARV-2 and LAV-1a have been reported elsewhere (Sanchez-Pescador et al., 1985; Wain-Hobson et al., 1985). Data for HAT-3 and WMJ-1 are from the present study. As discussed in the text and shown here, the entire LTR, NCR/*tat*, envelope, and 3' orf, as well as most of *gag*, *pol*, and *sor*, were sequenced for HAT-3. For WMJ-1, the entire envelope gene was sequenced.

^a For WMJ-2, a highly related viral clone from this same patient, the complete sequence of the *gag* p24 and p15 is shown (Hahn et al., 1986).

III/LAV transactivator (*tat* III) gene (Ratner et al., 1985a; Wain-Hobson et al., 1985; Sanchez-Pescador et al., 1985; Rosen et al., 1986). For HAT-3, we determined the nucleotide sequence of the entire LTR, *env*, 3' orf, and NCR (*tat*) regions as well as 84% of *gag*, 41% of *pol*, and 76% of *sor*. For WMJ-1, only the *env* gene was sequenced, but for two other WMJ clones that were closely related to WMJ-1, both *env* and *gag* (p15 and p24 regions) were sequenced (Hahn et al., 1986). From this analysis, and from restriction enzyme mapping and heteroduplexing, we confirmed that the overall genomic organization of HAT-3 and WMJ-1 is the same as that described previously for HTLV-III, LAV, and ARV.

Table 1 compares the nucleotide and deduced amino acid sequences of HAT-3, WMJ-1, LAV-1a, and ARV-2 to the original HTLV-IIIb clone, BH-10. These data demonstrate the existence of a broad spectrum of diversity among the AIDS retroviruses: Overall, BH-10 differed from HAT-3 in 8.6% of nucleotides (12.5% amino acids), from ARV-2 in 6.3% of nucleotides (9.2% amino acids), and from LAV-1a in 1.5% of nucleotides (2.2% amino acids). HAT-3 and ARV-2 differed from each other over the lengths of their sequenced genomes by 9.3% of nucleotides (14.2% amino acids). A similar spectrum of diversity among independent HTLV-III/LAV isolates has been reported on the basis of restriction enzyme mapping (Wong-Staal et al., 1985; Benn et al., 1985).

The distribution of sequence differences was not uniform throughout the viral genomes. Instead, when divergent genomes were compared, changes were found to be considerably more prevalent in the envelope and 3' orf than in other genes (Table 1). In the envelope gene in particular, BH-10 differed from HAT-3 in 12.9% of nucleotides (19% amino acids), from ARV-2 in 9.8% of nucleotides (15.7% amino acids), and from WMJ-1 in 9.3% of nucleotides (14.0% amino acids). In the same region, HAT-3 differed from ARV-2 in 14.7% of nucleotides (18.4% amino acids) and from WMJ-1 in 11.3% of nucleotides (16.5% amino acids). ARV-2 and WMJ-1 differed from each other in this region by 10.0% of nucleotides (16.2% amino acids). Within the envelope gene, the signal peptide and extracellular portion were the regions that were most highly divergent. For example, the extracellular envelope domains of HAT-3, WMJ-1, and ARV-2 differed from that of BH-10 in 21.4%, 18.7%, and 17.0% of amino acids, respectively. The two Haitian isolates, HAT-3 and WMJ-1, differed from each other in this region by 21.6% amino acids and from ARV-2 by 20.2% and 20.8%, respectively.

The most conserved areas in the genomes of HAT-3, LAV-1a, BH-10, and ARV-2 were the *gag* and *pol* genes, which differed among the viruses in less than 6% of nucleotides and less than 7% of amino acids. In these regions, nucleotide sequence changes were almost exclusively due to point mutations in contrast to *env* where

clustered nucleotide changes involving in-frame deletions, insertions, and/or duplications were common. For example, compared to BH-10, there were 48 bp of insertions and 21 bp of deletions in the extracellular envelope of HAT-3 as compared to only 6 bp of insertions and deletions in all of its analyzed *gag*, *pol*, and *sor* sequences. Similarly, there were 33 bp of insertions and 36 bp of deletions in the extracellular envelope of ARV-2 (compared to BH-10) but only a 6 bp insertion and no deletions in all of its *gag*, *pol*, and *sor*. Even in LAV-1a, which is more closely related to BH-10 than are the other sequenced viruses, there was a 15 bp insertion (duplication) in the exterior envelope gene but no insertions or deletions in *gag*, *pol*, or *sor*. Not included in this comparative analysis is a single direct repeat at the *gag/pol* junction of BH-10 that is not present in any of the other isolates.

Another difference in the types of mutations present in the *env* gene compared to the *gag* gene of these viruses was in the proportion of silent third base pair changes. Excluding deletions and insertions, compared to BH-10 the mutations in the external *env* genes of ARV-2, HAT-3, and WMJ (clone WMJ-2 from patient WMJ; see Hahn et al., 1986) were in the third position of the codon in 36%, 37%, and 34% of instances leading to 32%, 25% and 42% amino acid changes, respectively. Conversely, in the *gag* genes of these same viruses, 63%, 56%, and 66% of nucleotide changes were in the third place of the codon leading to 10%, 3.5% and 0% amino acid changes, respectively. Thus more than half of the single nucleotide changes in *env* occurred in the first or second codon position, and even for third position changes, many of these led to amino acid changes. Conversely, only a minority of first or second codon position changes occurred in *gag*, and in this gene the third position changes were almost uniformly silent. These findings are consistent with two interpretations; first, that structure-function relationships lead to stronger conservation of amino acid sequence in *gag* than in *env*, and, second, that non-silent nucleotide changes in *env* may actually be under positive selection pressure. Whether or not the latter interpretation is correct must await a much more complete understanding of the host immunologic response to HTLV-III/LAV.

Alignment and Comparison of Envelope Sequences

An alignment of the envelope nucleotide and amino acid sequences of HAT-3, WMJ-1, BH-10, LAV-1a, and ARV-2 is shown in Figures 1 and 2. A methionine codon at position 8 of the BH-10 envelope open reading frame most likely marks the beginning of the gene, with a hydrophobic region between position 17 and 37 representing the potential leader sequence. It is known from the amino acid sequence of the mature envelope glycoprotein that this leader sequence, or signal peptide, is cleaved from the envelope precursor protein during envelope maturation (Allen et al., 1985). The ensuing peptide sequence from position 38 to 518 (BH-10 sequence) is slightly hydrophilic and contains numerous potential N-linked glycosylation sites, ranging from 22 in WMJ-1 to 29 in HAT-3. This region corresponds to the major exterior envelope glycoprotein (gp120). A characteristic arginine-rich hydrophobic stretch

marks the cleavage site for the processing of the envelope precursor gp160 into the exterior gp120 and membrane-bound gp41 (Allen et al., 1985; DiMarzo-Veronese et al., 1985). The latter protein includes an apparent hydrophobic membrane-spanning segment, a hydrophilic anchor sequence, and an additional carboxy-terminal stretch of 150 residues whose function is presently unknown.

A striking feature of the five envelope sequences was the conservation of cysteine residues (Figure 2). Within the extracellular envelope, each of the 18 cysteine residues was conserved in all five viruses. Within the transmembrane protein, all but one cysteine residue in its extreme carboxy-terminus was conserved. This finding argues for a highly conserved "macrostructure" of different HTLV-III/LAV envelope glycoproteins. Despite this overall conservative structure, numerous amino acid changes were evident throughout the envelope glycoprotein. In the transmembrane region these were generally the result of isolated nucleotide point mutations leading to single amino acid substitutions. In the extracellular region, changes in amino acid sequence resulted from both nucleotide point mutations and in-frame deletions, insertions, or duplications of blocks of nucleotides. The resulting amino acid alterations appeared as clustered mutations interspersed with polypeptide segments that were highly conserved. The regions of highest variability (dark shading in Figure 2) and highest conservation (light shading in Figure 2) were identified both by visual inspection and by objective determination of relative variation using computer analysis (Figure 3).

Predicted Secondary Structure

Since the extracellular region of the HTLV-III/LAV envelope glycoprotein constitutes a major target of the host immunologic response (Barin et al., 1985a), we examined this region for predicted antigenic epitopes using a computer program that predicts the secondary structure of proteins superimposed with values for hydrophilicity (Chou and Fasman, 1974; Hopp and Woods, 1981). Such an analysis of other proteins, including viral envelopes, has shown that antigenic epitopes are often associated with hydrophilic protein domains containing β turns (Atassi, 1978; Cohen et al., 1984; Westhoff et al., 1984; Eisenberg et al., 1985; Gunn et al., 1985; Pellett et al., 1985). This analysis demonstrated that the exterior envelope proteins of the five AIDS retroviruses each contain a number of sites that meet criteria for likely antigenic epitopes and that these regions generally coincide with the variable regions identified independently by amino acid sequence comparisons. Figure 4 illustrates the substantial differences in predicted secondary protein structure for the first variable region of the external envelope gene product for the five viruses. In this region, there were 7 β turns in HAT-3, 0 β turns in WMJ-1, 8 β turns in BH-10, 11 β turns in LAV-1a, and 5 β turns in ARV-2. The five envelope sequences also differed in this region in their degrees of hydrophilicity and in the number and location of potential N-linked glycosylation sites (Figures 2 and 4). In each of the other variable regions substantial differences existed similarly in predicted hydrophilicity, second-

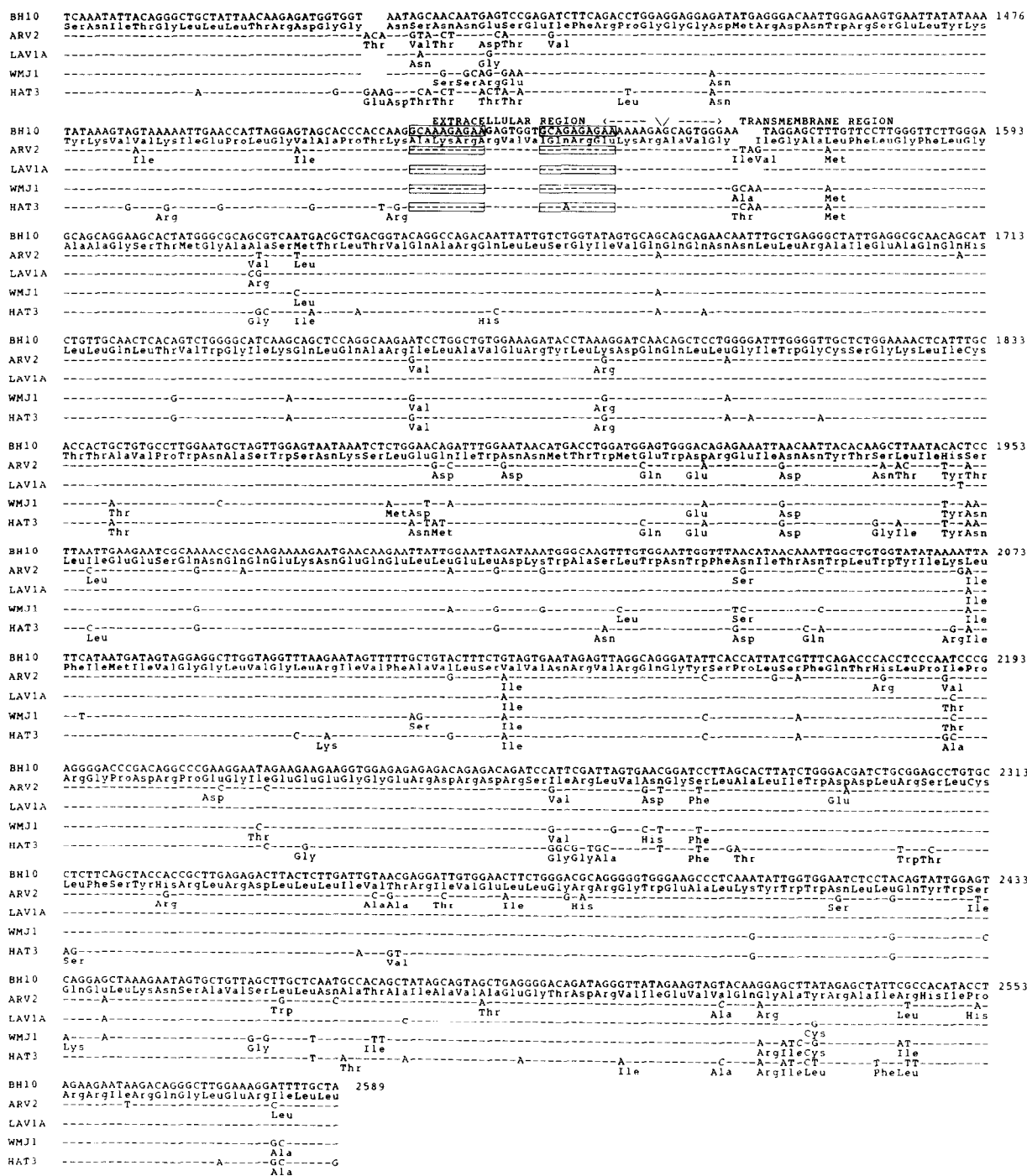


Figure 1. Nucleotide Sequences of the Entire Envelope Genes of Five Independent AIDS Virus Isolates

Sequence information for BH-10 (HTLV-IIIb), ARV-2, and LAV-1a was obtained from GENBANK. Sequences for WMJ-1 and HAT-3 were determined as part of the present study. Alignment of the sequences was performed pairwise with the assistance of NUCALN (Wilbur and Lipman, 1983). Numbering of nucleotides is from the first nucleotide of BH-10 following the TAG stop codon, and genomic regions corresponding to the signal peptide, extracellular glycoprotein (gp120), and transmembrane glycoprotein (gp 41) are shown. Dashes indicate nucleotide identity with BH-10 and spaces indicate the absence of that nucleotide. The boxed nucleotides comprise adjacent perfect or imperfect direct repeats.

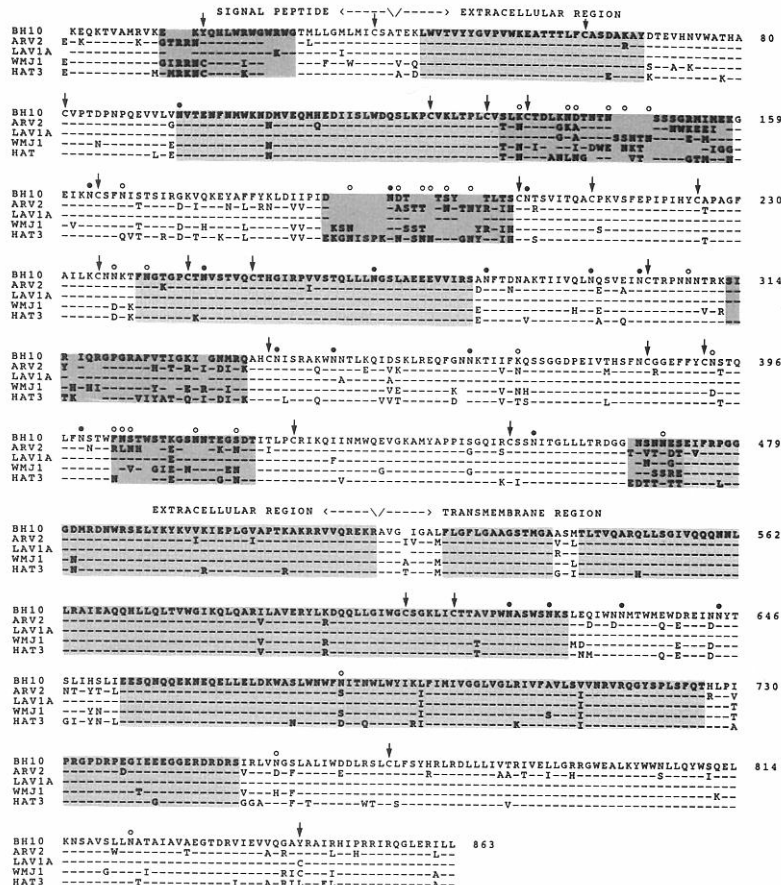


Figure 2. Amino Acid Sequences of the Entire Envelope Genes of Five Independent AIDS Virus Isolates

Sequence information for BH-10 (HTLV-IIIb), ARV-2, and LAV-1a was obtained from GENBANK and information for WMJ-1 and HAT-3 from the present study. Alignment of the sequences was performed pairwise with the assistance of PRTALN (Wilbur and Lipman, 1983). Numbering of amino acids is from the first amino acid of BH-10 and regions corresponding to the signal peptide, extracellular envelope glycoprotein (gp120), and transmembrane glycoprotein (gp41) are shown. Dashes indicate amino acid identity with BH-10 and spaces indicate the absence of that amino acid. Arrows denote cysteine residues and solid and open circles denote conserved and nonconserved sites of potential N-linked glycosylation, respectively. Darkly shaded regions within the extracellular envelope glycoprotein correspond to regions of hyper-variability as determined by both visual inspection and by computer analysis of variation (see Figure 3). Lightly shaded regions correspond to areas that are relatively highly conserved. Note that additional regions of intermediate variability are interspersed among highly variable and highly conserved areas.

ary structure, and potential glycosylation sites (S. Modrow, unpublished data).

Interspersed with the variable regions of the exterior envelope protein were other areas that were highly conserved among all five isolates analyzed. Most of these regions were primarily hydrophobic and contained only few or no β turns. Accordingly, they are believed to be less likely to represent antigenic epitopes. There were, however, a number of exceptions. One such exception is a conserved stretch of 46 amino acid residues immediately adjacent to the processing site of the envelope precursor (amino acids 473 through 518 in Figure 2). This area, shown also in Figure 4, was very highly conserved among all five viruses, contained numerous β turns, and was

hydrophilic. Thus this conserved region of the exterior envelope glycoprotein would be expected to be both antigenic and cross-reactive among different viral strains. It has recently been shown that a synthetic peptide generated from a portion of this region does in fact detect naturally occurring antibodies in many patients infected with HTLV-III/LAV and that monoclonal antibodies directed against peptides of this region recognize the native gp120 (T. Palker and B. Haynes, personal communication).

Discussion

A growing body of scientific evidence indicates that the rate of evolution for RNA viruses in general is much

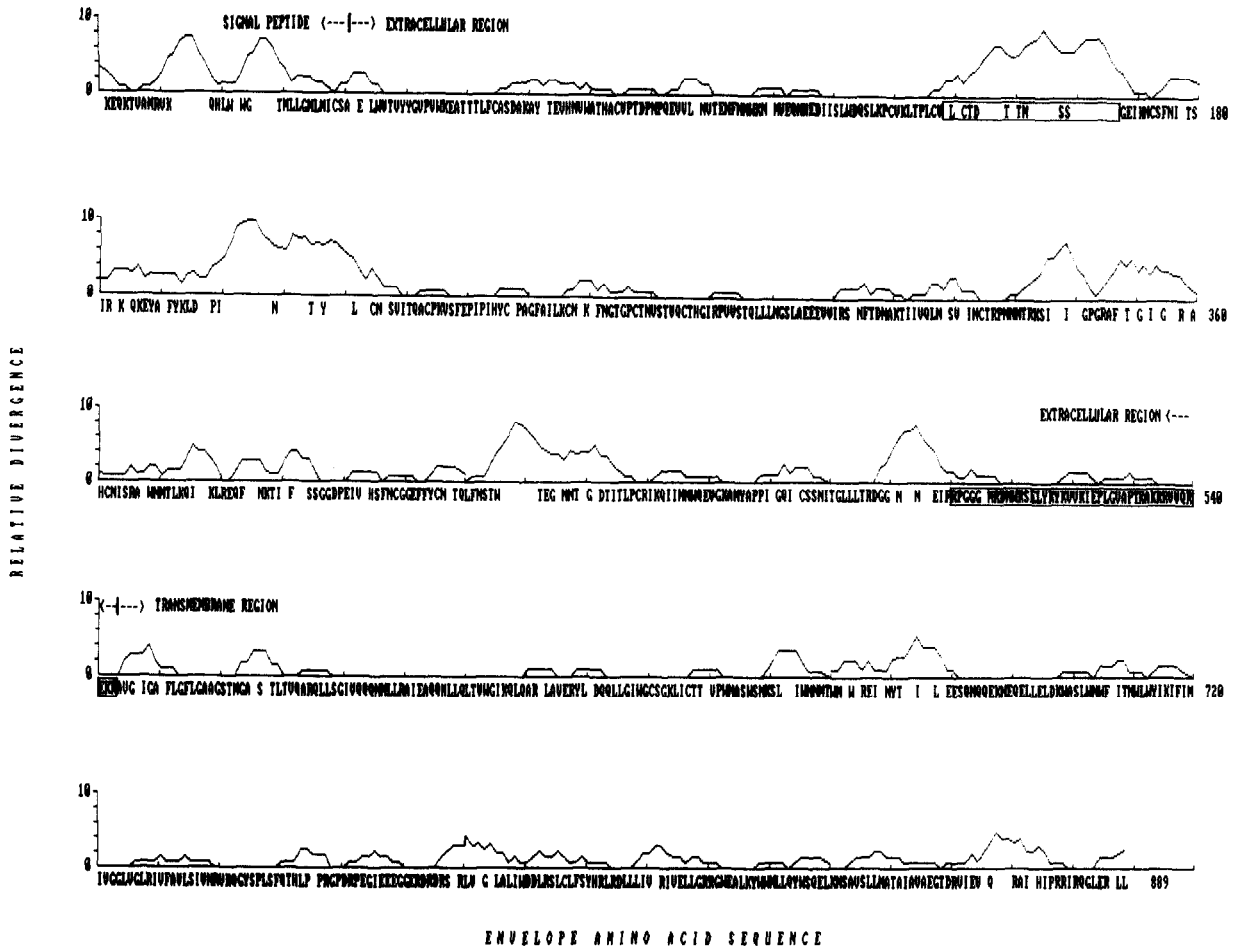


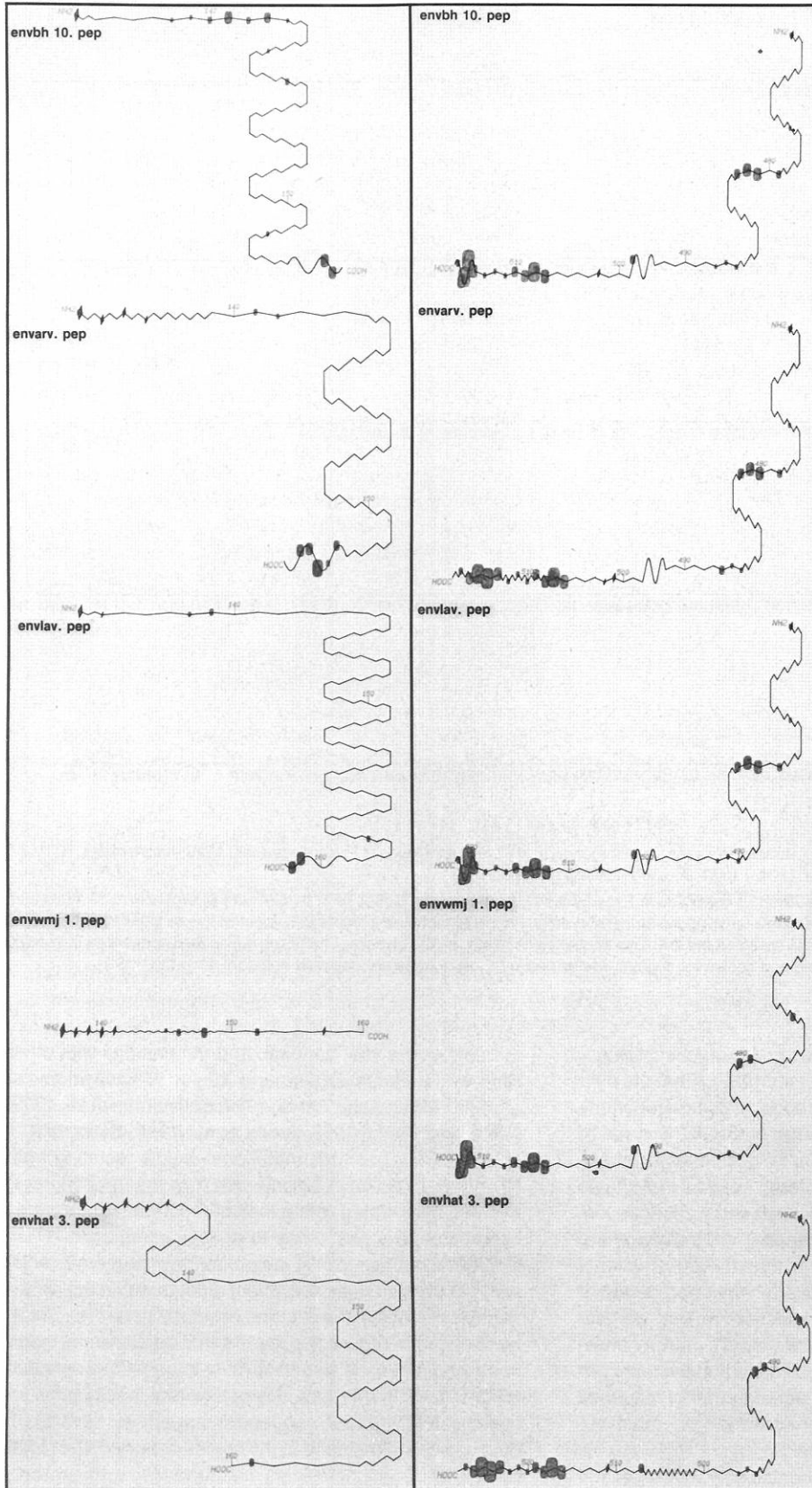
Figure 3. Computer Graphic Illustrating Relative Variation, or Divergence, in the Envelope Genes of Five Independent AIDS Virus Isolates, HTLV-IIIb (BH-10), LAV (LAV-1a), ARV (ARV-2), HTLV-III_{RF} (HAT-3), and HTLV-III_{WJ-1} (WMJ-1)

Relative degrees of variation were calculated as described in Experimental Procedures and plotted from 0 (minimum divergence) to 10 (maximum divergence). Beneath the abscissa are single-letter codes for conserved amino acids that were present in five out of five or four out of five of the sequences. Numbering of amino acids is based on the envelope sequence as shown on the abscissa. The boxed amino acid sequences in line 180 and line 540 correspond to the variable and conserved regions whose secondary structures are depicted in Figure 4.

greater than for most prokaryotic or eukaryotic DNA genomes (for review, see Holland et al., 1982). From the data reported here, it is apparent that substantial heterogeneity exists among independent isolates of the AIDS virus. In another study (Hahn et al., 1986), we examined serial isolates of HTLV-III/LAV from chronically infected individuals and showed that the types of nucleotide changes described in this paper for independent virus isolates are also present in serial virus isolates from the same patients and that these changes occur over a very short period of time during the course of viral infection. In fact, we estimate the rate of genetic change for the HTLV-III/LAV envelope gene to be at least 10^{-3} nucleotide substitutions per site per year, a mutational rate equal to that of influenza A virus and a million times greater than for most eukaryotic genes (Hahn et al., 1986).

An understanding of the mechanisms by which genetic variation occurs in HTLV-III/LAV can be approached by comparative analysis of the five sequenced genomes described herein. The finding of frequent point mutations

throughout the viral genome, and the analogy with other RNA viral systems (Zarling and Temin, 1975; Clements et al., 1980; O'Rear and Temin, 1982; Shivelman et al., 1983; Darlix and Spahr, 1983; Bruck et al., 1984; Salinovich et al., 1986), suggest that mutational events resulting from RNA-dependent DNA polymerase (reverse transcriptase) misreading coupled with a lack of proofreading enzymes contributes in a major way to genetic variation in HTLV-III/LAV. In addition to this, the presence of numerous adjacent perfect and imperfect direct repeats within the hyper-variable regions of the external envelope gene (see boxed nucleotides in Figure 1) suggests that duplications, insertions, and deletions also contribute to genetic variation in HTLV-III/LAV. In this light, it is noteworthy that in some instances HTLV-III/LAV sequences appear to have been duplicated and then one of the copies changed, the result being that novel amino acids are introduced within that region. Direct repeats encompassing deletions have been observed in other biologic systems, both eukaryotic (Efstratiadis et al., 1980) and prokaryotic (Farabaugh et al.,



1978), and in the *lac I* gene of *E. coli* such direct repeats correspond to hotspots of spontaneous genetic mutation (Farabaugh et al., 1978).

Other potential mechanisms for genomic change in HTLV-III/LAV include recombination between different viral DNA molecules or between viral DNA and host DNA. Both processes are known to occur in other retroviral systems (Coffin, 1979; O'Rear and Temin, 1982; Dorner et al., 1985; Bishop and Varmus, 1985). In this study, and in a related study in which serial AIDS virus isolates from three patients were analyzed by restriction mapping or DNA sequencing (Hahn et al., 1986), evidence for recombination between different viral molecules was not apparent. However, given the high rate of genetic change in HTLV-III/LAV and the persistence within a given individual of more than one predominant viral form (Shaw et al., 1984; Wong-Staal et al., 1985; Hahn et al., 1986), genetic recombination between different HTLV-III/LAV molecules is a distinct possibility. The second process, recombination between viral and cellular DNA, appears much less likely. None of the AIDS virus isolates studied here contained DNA sequences closely related to normal human DNA as determined by Southern blot hybridization (Hahn et al., 1984; Luciw et al., 1984; Alizon et al., 1984), and comparison of their envelope nucleotide sequences did not reveal any large unexplained insertions or rearrangements (see Figure 1). Still another potential source for genetic variation in HTLV-III/LAV is the formation of stable secondary structures (stem-loops) in the viral RNA which in turn could lead to localized areas of reverse transcriptase misreading. This process has been proposed to explain in part the variation observed in foot and mouth disease virus (Weddell et al., 1985), but for HTLV-III/LAV we found no evidence for such stem-loop structures. In sum, the nucleotide sequence data from independent HTLV-III/LAV isolates, along with the analysis of sequential viral isolates from individual patients (Hahn et al., 1986), suggest strongly that point mutations, in combination with short deletions or insertions/duplications perhaps related to copy-choice misreading by the viral polymerase, are the primary mechanisms by which variation is generated in the AIDS virus.

The finding of genomic variation primarily in the extracellular envelope gene in regions having properties predictive of antigenicity raises the possibility that host-related immunologic pressures may select for variant viral strains. For two retroviruses related to HTLV-III/LAV, namely equine infectious anemia virus (EIAV) and visna virus, there is evidence that progressive changes in the envelope genes do in fact lead to substantial changes in

envelope antigenicity (Clements et al., 1980; Montelaro et al., 1984; Salinovich et al., 1986). These genetic changes are localized predominantly within the envelope gene, in contrast to *gag* and *pol*, and there is evidence that they result from immunologic selective pressures exerted by the host. For EIAV, there is considerable support for the idea that such changes in viral antigenicity are directly responsible for the chronic, periodic nature of the disease (Salinovich et al., 1986). For visna virus, similar changes in envelope antigenicity leading to loss of virus neutralization by host antibodies have been noted (Clements et al., 1980). However, the biologic relevance of these findings has been questioned by other workers who have discovered spread of virus in blood and CSF unabated by neutralizing antibodies (Petursson et al., 1976), long-term persistence of parental viral strains even after the appearance of variants (Lutley et al., 1983), and lack of antigenic variants in some animals even during advanced stages of disease (Thormar et al., 1983). Given the complexities of chronic retroviral-induced disease (Haase et al., 1977; Brahic et al., 1981; Stowring et al., 1985), it is likely that intrinsic viral characteristics as well as host-viral immune interactions play important roles in clinical outcome.

For HTLV-III/LAV, the biologic significance of envelope variation has been particularly difficult to ascertain because the nature of the host immunologic response to the virus has yet to be well characterized. Naturally occurring, cross-reactive, neutralizing antibodies specific for HTLV-III/LAV have been identified in individuals infected with HTLV-III/LAV (Robert-Guroff et al., 1985; Weiss et al., 1985). Furthermore, preliminary data also indicate that type-specific anti-HTLV-III/LAV antibodies are generated in vivo and that variants of HTLV-III/LAV evolve which, determined by their susceptibility to neutralization by homologous sera are antigenically distinct (M. Robert-Guroff and W. Parks, personal communication). Whether or not such type-specific neutralizing antibodies select for biologically important antigenic variants of HTLV-III/LAV during persistent viral infection in man remains to be seen.

Other mechanisms, unrelated to the immune system, could conceivably be involved in the selection of HTLV-III/LAV variants. For example, host-specific and tissue-specific properties of the cell membrane receptor for the virus could select for alterations in the viral envelope as the virus spreads through genetically different individuals or through different tissues within a given host (Shaw et al., 1985; Ho et al., 1985). Such a mechanism has been proposed as a possible explanation for envelope divergence in foot and mouth disease virus (FMDV), since this virus naturally infects many different animal species in-

Figure 4. Predicted Secondary Structure and Hydrophilicity Patterns of Representative Variable and Conserved Envelope Regions of BH-10, ARV-2, LAV-1a, WMJ-1, and HAT-3

The left-hand panel corresponds to the first major variable region of the extracellular viral envelope, amino acids 137-158 in Figure 2. The right-hand panel corresponds to the largest conserved region, which, in addition, possesses structural and hydrophilicity properties suggestive of antigenicity. This region corresponds to amino acids 473-518 immediately 5' to the envelope processing site shown in Figure 2. Ovals denote hydrophilic regions and diamonds hydrophobic regions. The radius of an oval or a diamond over a residue is proportional to the mean hydrophilicity or hydrophobicity as calculated for that residue plus the next four residues. Positions of β turns are indicated by turns of the chain (\equiv), α helices as bold coils ($\sim\sim\sim$), β sheets as narrow zigzags ($\sim\sim\sim$), and random coils as wide zigzags ($\sim\sim\sim$).

cluding cattle, swine, sheep, goats, buffalo, and antelope (Weddell et al., 1985). A selection process for envelope variation related to cell membrane characteristics within and between species is also of special relevance to HTLV-III/LAV research, as are mechanisms involving genetic recombination, since a novel simian retrovirus (STLV-III, for Simian T-Lymphotropic Virus Type III) recently isolated from four diseased rhesus macaques (three with immunodeficiency and one with transmitted lymphoma) has been found to be very similar to HTLV-III/LAV in its morphology, protein structure, and immunologic characteristics (Kanki et al., 1985; Daniel et al., 1985). Antibodies to this virus (or a close relative) were identified in 28 of 67 healthy African green monkeys from central Africa (caught in the wild), as well as in a number of clinically asymptomatic West African prostitutes (Barin et al., 1985b). The fact that this retrovirus has infected two species of Old World primates, and possibly man, and the fact that it is highly related to HTLV-III/LAV suggests that the evolutionary origins of STLV-III and HTLV-III/LAV are similar. That their envelope proteins are less related than their *gag* proteins underscores the potential importance of variation in the envelope gene of HTLV-III/LAV as a determinant of biologic activity.

Whether or not the observed propensity for variation in HTLV-III/LAV has biologic significance has yet to be formally proven. However, such extensive variation in the different genes of this virus clearly has the potential for drastically altering both its immunologic and biologic properties. These include the antigenicity and tissue tropism conferred by *env*; the polymerase, protease, and integrase functions of *pol*; the gene regulatory properties of LTR and *tat*; and other potentially important products of *src* and 3' *orf* whose functions are still unknown. Future efforts to understand viral pathogenesis and to develop effective preventive and treatment measures for AIDS will have to take into account the extensive genomic heterogeneity present in HTLV-III/LAV.

Experimental Procedures

Virus Isolation

Virus was isolated from patients' peripheral blood mononuclear cells, transmitted to immortalized T cell lines, and propagated in tissue culture as described (Popovic et al., 1984; Hahn et al., 1986).

Molecular Cloning of HAT-3 and WMJ-1

Lambda (λ) phage libraries were constructed according to standard procedures (Maniatis et al., 1982) using the cloning vector λ gt Wes- λ B and SstI digested viral/cellular DNA which had been enriched by sucrose gradient centrifugation for 8–10 kb fragments. Both HTLV-III_{RF} and HTLV-III_{WMJ-1} DNA contained a single 9 kb SstI fragment detectable by Southern hybridization which corresponds to almost the entire HTLV-III/LAV provirus in nonpermuted form. Phage plaques (approximately 5×10^5) were screened with the HTLV-III/LAV-specific 9 kb SstI viral insert from λ BH-10 (Hahn et al., 1984) that had been subcloned into SP64. Positive signals were plaque-purified and characterized. For nucleotide sequencing, the 9 kb inserts from both HAT-3 (a clone from isolate HTLV-III_{RF}) and WMJ-1 (a clone from isolate HTLV-III_{WMJ-1}) were self-ligated, cut with PstI, and subcloned into the PstI site of either pBR322 or SP64. Sequencing reactions were then performed using these subgenomic clones. Detailed restriction cleavage patterns of both HAT-3 and WMJ-1 corresponded exactly (25 out of 25 restriction sites) to the predominant proviral forms present in the respective cell lines.

Nucleotide Sequence Analysis

Nucleotide sequencing was performed according to Maxam and Gilbert (1980). Plasmid subclones were cleaved with the appropriate restriction enzymes and end-labeled using polynucleotide kinase and γ - 32 P-ATP for 5' ends, and α - 32 P-dNTP plus DNA polymerase (Klenow fragment) or α - 32 P-ddATP plus terminal deoxyribonucleotidyl transferase for 3' ends. Labeled fragments were cleaved with a second enzyme, separated by polyacrylamide gel electrophoresis, and eluted into buffer. In some instances, double-labeled fragments were prepared, strand-separated, and sequenced individually. The sequences of both envelope genes and most of the remainder of λ HAT-3 was confirmed by sequence analysis of both DNA strands. The computer program of Queen and Korn (1980) was used for translating the nucleotide sequences into amino acids. Nucleotide and amino acid sequences were aligned pairwise using NUCALN and PRTALN (Wilbur and Lipman, 1983).

Computer-Assisted Analysis of Envelope Variation

A computer program was written that compared the aligned amino acid sequences of BH-10, ARV-2, LAV-1A, WMJ-1, and HAT-3 and plotted relative variation, or divergence, as a function of amino acid position within the envelope. For this analysis, the aligned amino acid sequences shown in Figure 2 were analyzed column by column. Column scores were computed based on the number of mismatched pairs with values ranging from 0 (no mismatches, complete homology) to 10 (all mismatches, no homology). The absence of an amino acid (deletion) at any position was considered to be a mismatch. Each column score was averaged over two column scores on each side resulting in a window size of five. This mean value for each amino acid position was plotted from 0 to 10 above the corresponding envelope amino acid position shown on the abscissa.

Secondary Structure Analysis

A computer program that predicts secondary structures of proteins according to the rules of Chou and Fasman (1974) and superimposes values for hydrophilicity according to Hopp and Woods (1981) was employed to analyze the envelope amino acid sequences of the five different AIDS virus clones. This program, originally designed by E. Golub (Cohen et al., 1984), was modified slightly and adapted to the VAX750 computer (S. Modrow, unpublished data). The probability of the occurrence of α helices, β pleated sheets, random coils, and β turn regions were evaluated using stringent conditions: p boundary ≥ 1 , with $p\beta > p\alpha$ and $p_t > p_\alpha$ ($p\beta$ = probability for β turn region). The parameters for hydrophilicity were averaged over five amino acid residues with a limit of 0.7. β turns adjacent to β sheets or α helical regions in a non-hydrophobic environment are likely candidates for antigenic sites since they frequently form loop-like structures on the outside of the protein surface (Westhoff et al., 1984).

Acknowledgments

We thank Ms. Sheila Chastain for careful preparation of the manuscript. This work was supported in part by grants 1 U01 AI 23616-01, P30 CA13148, AI 21122, and AI 20736 from the National Institutes of Health. B. H. H. is a Special Fellow of the Leukemia Society of America. G. M. S. is a Pew Scholar in the Biomedical Sciences.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received July 17, 1985; revised April 29, 1986.

References

- Alizon, M., Sonigo, P., Barré-Sinoussi, F., Chermann, J.-C., Tiollais, P., Montagnier, L., and Wain-Hobson, S. (1984). Molecular cloning of lymphadenopathy-associated virus. *Nature* 312, 757–760.
- Allen, J. S., Coligan, J. E., Barin, F., McLane, M. F., Sodroski, J. G., Rosen, C. A., Haseltine, W. A., Lee, H. T., and Essex, M. (1985). Major glycoprotein antigens that induce antibodies in AIDS patients are encoded by HTLV-III. *Science* 228, 1091–1093.
- Atassi, M. Z. (1978). Precise prediction of the entire antigenic structure

- of lysozyme. Molecular features of protein antigenic structures and potential of "surface stimulation" synthesis—a powerful new concept for protein binding sites. *Immunochemistry* 15, 909–936.
- Barin, F., McLane, M. F., Allan, J. S., Lee, T. H., Groopman, J. E., and Essex, M. (1985a). Virus envelope protein of HTLV-III represents major target antigen for antibodies in AIDS patients. *Science* 228, 1094–1096.
- Barin, F., Denis, F., Allan, J. S., M'Boup, S., Kanki, P., Lee, T. H., and Essex, M. (1985b). Serological evidence for virus related to simian T-lymphotropic retrovirus III in residents of West Africa. *Lancet* ii, 1387–1389.
- Barré-Sinoussi, F., Chermann, J. C., Rey, F., Nugeyre, M. T., Chamaret, S., Gruest, J., Danguet, C., Axler-Blin, C., Vézinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L. (1983). Isolation of T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* 220, 868–870.
- Benn, S., Rutledge, R., Folks, T., Gold, J., Baker, L., McCormick, J., Feorino, P., Piot, P., Quinn, T., and Martin, M. A. (1985). Genomic heterogeneity of AIDS retroviral isolates from North America and Zaire. *Science* 230, 949–951.
- Bishop, J. M., and Varmus, H. (1985). Functions and origins of retroviral transforming genes. In *RNA Tumor Viruses*. Ed. by R. Weiss, N. Teich, H. Varmus, and J. Coffin (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 301–327.
- Brahic, M., Stowring, L., Ventura, P., and Haase, A. T. (1981). Gene expression in visna virus infection in sheep. *Nature* 292, 240–242.
- Bruck, C., Rensonnet, N., Portetelle, D., Cleuter, Y., Mammerickx, M., Burny, A., Mamoun, R., Guillemain, B., Van Der Maaten, M. J., and Ghysdael, J. (1984). Biologically active epitopes of bovine leukemia virus glycoprotein gp51: their dependence on protein glycosylation and genetic variability. *Virology* 136, 20–31.
- Chou, P. Y., and Fasman, G. D. (1974). Prediction of protein conformation. *Biochemistry* 13, 222–245.
- Clements, J. E., Pedersen, F. S., Narayan, O., and Haseltine, W. A. (1980). Genomic changes associated with antigenic variation of visna virus during persistent infection. *Proc. Natl. Acad. Sci. USA* 77, 4454–4458.
- Coffin, J. M. (1979). Structure, replication, and recombination of retrovirus genomes: some unifying hypotheses. *J. Gen. Virol.* 42, 1–46.
- Cohen, G. H., Dietzschold, B., Ponce de Leon, M., Long, D., Golub, E., Varrichio, A., Pereira, L., and Eisenberg, R. J. (1984). Localization and synthesis of an antigenic determinant of Herpes simplex virus glycoprotein D that stimulates the production of neutralizing antibody. *J. Virol.* 49, 102–108.
- Daniel, M. D., Letvin, N. L., King, N. W., Kannagi, M., Sehgal, P. K., Hunt, R. D., Kanki, P. J., Essex, M., and Desrosiers, R. C. (1985). Isolation of T-cell tropic HTLV-III-like retrovirus from macaques. *Science* 228, 1201–1204.
- Darlix, J.-L., and Spahr, P.-F. (1983). High spontaneous mutation rate of Rous sarcoma virus demonstrated by direct sequencing of the RNA genome. *Nucl. Acid Res.* 11, 5953–5967.
- DiMarzo-Veronese, F., deVico, A. L., Copeland, T. D., Oroszlan, S., Gallo, R. C., and Sarngadharan, M. G. (1985). Characterization of gp41 as the transmembrane protein coded by the HTLV-III/LAV envelope gene. *Science* 229, 1402–1405.
- Dorner, A. J., Stoye, J. P., and Coffin, J. M. (1985). Molecular basis of host range variation in avian retroviruses. *J. Virol.* 53, 32–39.
- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, R. M., O'Connell, C., Spritz, R. A., DeRiel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C., and Proudfoot, N. J. (1980). The structure and evolution of the human β -globin gene family. *Cell* 21, 653–668.
- Eisenberg, R. J., Long, D., Ponce de Leon, M., Matthews, J. T., Spear, P. G., Gibson, M. G., Lasky, L. A., Berman, P., Golub, E., and Cohen, G. H. (1985). Localization of epitopes of Herpes simplex virus type I glycoprotein D. *J. Virol.* 53, 634–644.
- Farabaugh, P. J., Schmeissner, U., Hofer, M., and Miller, J. H. (1978). Genetic studies of the lac repressor VII. On the molecular nature of spontaneous hotspots in the lacI gene of *Escherichia coli*. *J. Mol. Biol.* 126, 847–863.
- Gallo, R. C., Salahuddin, S. Z., Popovic, M., Shearer, G. M., Kaplan, M., Haynes, B. F., Palker, T. J., Redfield, R., Oleske, J., Safai, B., White, G., Foster, P., and Markham, P. D. (1984). Human T-lymphotropic retrovirus, HTLV-III, isolated from AIDS patients and donors at risk for AIDS. *Science* 224, 500–503.
- Gunn, P. R., Sato, F., Powell, K. F. H., Bellamy, A. R., Napier, J. R., Harding, D. R. K., Hancock, W. S., Siegman, L. J., and Both, G. W. (1985). Rotavirus neutralizing protein VP7: antigenic determinants investigated by sequence analysis and peptide synthesis. *J. Virol.* 54, 791–797.
- Haase, A., Stowring, L., Narayan, O., Griffin, D., and Price, D. (1977). Slow persistent infection caused by visna virus: role of host restriction. *Science* 195, 175–177.
- Hahn, B., Shaw, G. M., Arya, A. K., Popovic, M., Gallo, R. C., and Wong-Staal, F. (1984). Molecular cloning and characterization of the virus associated with AIDS (HTLV-III). *Nature* 312, 166–169.
- Hahn, B. H., Gonda, M. A., Shaw, G. M., Popovic, M., Hoxie, J., Gallo, R. C., and Wong-Staal, F. (1985). Genomic diversity of the AIDS virus HTLV-III: different viruses exhibit greatest divergence in their envelope genes. *Proc. Natl. Acad. Sci. USA* 82, 4813–4817.
- Hahn, B. H., Shaw, G. M., Taylor, M. E., Redfield, R. R., Markham, P. D., Salahuddin, S. Z., Wong-Staal, F., Gallo, R. C., Parks, E. S., and Parks, W. P. (1986). Genetic variation in HTLV-III/LAV over time in patients with AIDS or at risk for AIDS. *Science*, in press.
- Ho, D. D., Rota, T. R., Schooley, R. T., Kaplan, J. C., Allan, J. D., Groopman, J. E., Resnick, L., Felsenstein, D., Andrews, C. A., and Hirsch, M. S. (1985). Isolation of HTLV-III from cerebrospinal fluid and neural tissues of patients with neurologic syndromes related to the acquired immunodeficiency syndrome. *N. Engl. J. Med.* 313, 1493–1497.
- Holland, J., Spindler, K., Horodyski, F., Grabau, E., Nichol, S., and Vande Pol, S. (1982). Rapid evolution of RNA genomes. *Science* 215, 1577–1585.
- Hopp, T. P., and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* 78, 3824–3828.
- Kanki, P. J., McLane, M. F., King, N. W., Letvin, N. L., Hunt, R. D., Sehgal, P., Daniel, M. D., Desrosiers, R. C., and Essex, M. (1985). Serologic identification and characterization of a macaque T-lymphotropic retrovirus closely related to HTLV-III. *Science* 228, 1199–1201.
- Luciw, P. A., Potter, S. J., Steimer, K., Dina, D., and Levy, J. A. (1984). Molecular cloning of AIDS-associated retrovirus. *Nature* 312, 760–763.
- Lutley, R., Petursson, C., Pálsson, P. A., Georgsson, G., Klein, J., and Nathanson, N. (1983). Antigenic drift in visna: virus variation during long-term infection of Icelandic sheep. *J. Gen. Virol.* 64, 1433–1440.
- Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982). *Molecular Cloning* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 326–328.
- Maxam, A. M., and Gilbert, W. (1980). Sequencing end-labeled DNA with base-specific chemical cleavages. *Meth. Enzymol.* 65, 499–560.
- Montelaro, R. C., Parekh, B., Orrego, A., and Issel, C. J. (1984). Antigenic variation during persistent infection by equine infectious anemia virus, a retrovirus. *J. Biol. Chem.* 259, 10539–10544.
- Muesing, M. A., Smith, D. H., Cabradilla, C. D., Benton, C. V., Lasky, L. A., and Capon, D. J. (1985). Nucleic acid structures and expression of the human AIDS/lymphadenopathy retrovirus. *Nature* 313, 430–458.
- O'Rear, J. J., and Temin, H. M. (1982). Spontaneous changes in nucleotide sequence in proviruses of spleen necrosis virus, an avian retrovirus. *Proc. Natl. Acad. Sci. USA* 79, 1230–1234.
- Pellett, P. E., Konsonlas, K. G., Pereira, L., and Roizman, B. (1985). Anatomy of Herpes simplex virus 1 strain F glycoprotein B gene: primary sequence and predicted protein structure of the wild type and of monoclonal antibody-resistant mutants. *J. Virol.* 53, 243–253.
- Petursson, G., Nathanson, N., Georgsson, G., Panitch, H., and Pálsson, P. A. (1976). Pathogenesis of visna, sequential virologic, serologic, and pathologic studies. *Lab. Invest.* 35, 402–412.
- Popovic, M., Sarngadharan, M. G., Read, E., and Gallo, R. C. (1984). A method for detection, isolation, and continuous production of cytopathic human T-lymphotropic retroviruses of the HTLV family (HTLV-III) from patients with AIDS and pre-AIDS. *Science* 224, 497–500.

- Queen, C. L., and Korn, L. J. (1980). Computer analysis of nucleic acids and proteins. *Meth. Enzymol.* 65, 595-609.
- Rabson, A., and Martin, M. A. (1985). Molecular organization of the AIDS retrovirus. *Cell* 40, 477-480.
- Ratner, L., Haseltine, W., Patarca, R., Livak, K. J., Starcich, B., Josephs, S. F., Doran, E. R., Rafalski, J. A., Whitwhoen, E. A., Baummeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, F. (1985a). Complete nucleotide sequence of the AIDS virus HTLV-III. *Nature* 313, 277-284.
- Ratner, L., Gallo, R. C., and Wong-Staal, F. (1985b). HTLV-III, LAV, and ARV are variants of the same AIDS virus. *Nature* 313, 636-637.
- Robert-Guroff, M., Brown, M., and Gallo, R. C. (1985). HTLV-neutralizing antibodies in AIDS and ARV. *Nature* 316, 72-74.
- Rosen, C. A., Sodroski, J. G., Goh, W. C., Dayton, A. I., Lippke, J., and Haseltine, W. A. (1986). Post-transcriptional regulation accounts for the trans-activation of the human T-lymphotropic virus type III. *Nature* 319, 555-559.
- Salinovich, O., Payne, S. L., Montelaro, R. C., Hussain, K. A., Issel, C. J., and Schnorr, K. L. (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., Power, M. D., Barr, P. J., Steimer, K. S., Stempfen, M. M., Brown-Shimer, S. L., Gee, W. W., Renard, A., Randolph, A., Levy, J. A., Dina, D., and Luciw, P. A. (1985). Nucleotide sequence and expression of an AIDS-associated retrovirus (ARV-2). *Science* 227, 484-492.
- Sarngadharan, M., Popovic, M., Bruch, L., Schupbach, J., and Gallo, R. C. (1984). Antibodies reactive with human T-lymphotropic retrovirus (HTLV-III) in the sera of patients with acquired immune deficiency syndrome. *Science* 223, 506-508.
- Schupbach, J., Popovic, M., Gilden, R. V., Gonda, M. A., Sarngadharan, M. G., and Gallo, R. C. (1984). Serologic analysis of a new type of human T-lymphotropic retrovirus (HTLV-III) associated with AIDS. *Science* 224, 503-506.
- Shaw, G. M., Hahn, B. H., Arya, S. K., Groopman, J. E., Gallo, R. C., and Wong-Staal, F. (1984). Molecular characterization of human T-cell leukemia (lymphotropic) virus type III in the acquired immune deficiency syndrome. *Science* 226, 1165-1171.
- Shaw, G. M., Harper, M. E., Hahn, B. H., Epstein, L. G., Gaidusek, C. D., Price, R. W., Navia, B. A., Petit, C. K., O'Hara, C. J., Cho, E.-S., Oleske, J. M., Wong-Staal, F., and Gallo, R. C. (1985). HTLV-III infection in brains of children and adults with AIDS encephalopathy. *Science* 227, 177-182.
- Shtivelman, E., Zakut, R., and Canaani, E. (1983). Frequent generation of nonrescuable reorganized Moloney murine sarcoma viral genomes. *Proc. Natl. Acad. Sci. USA* 81, 294-298.
- Stowring, L., Haase, A. T., Petursson, G., Georgsson, G., Palson, P., Lutley, R., Roos, R., and Szuchet, S. (1985). Detection of visna virus antigens and RNA in glial cells in foci of demyelination. *Virology* 141, 311-318.
- Thormar, H., Barshatzky, M. R., Arnesen, K., and Kozlowski, P. B. (1983). The emergence of antigenic variants is a rare event in long-term visna virus infection in vivo. *J. Gen. Virol.* 64, 1427-1432.
- Wain-Hobson, S., Sonigo, P., Danos, O., Cole, S., and Alizon, M. (1985). Nucleotide sequence of the AIDS virus, LAV. *Cell* 40, 9-17.
- Weddell, G. N., Yansuta, D. G., Dowbenko, D. J., Hoatlin, M. E., Grubman, M. J., Moore, D. M., and Kleid, D. G. (1985). Sequence variation in the gene for the immunogenic capsid protein VP1 of foot-and-mouth disease virus type A. *Proc. Natl. Acad. Sci. USA* 82, 2618-2622.
- Weiss, R. A., Chapham, P. R., Cheingson-Popov, R., Dagleish, A. G., Carne, C. A., Weller, I. V. D., and Tedder, R. S. (1985). Neutralizing antibodies to human T-cell lymphotropic virus type III. *Nature* 316, 69-72.
- Westhoff, E., Altschuh, D., Moras, D., Bloomer, A. C., Mondragon, A., Klug, A., and van Regenmortel, M. H. V. (1984). Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature* 311, 123-126.
- Wilbur, W. J., and Lipman, D. J. (1983). Rapid similarity searches of nucleic acid and protein data banks. *Proc. Natl. Acad. Sci. USA* 80, 726-730.
- Wong-Staal, F., Shaw, G. M., Hahn, B. H., Salahuddin, S. Z., Popovic, M., Markham, P. D., Redfield, R., and Gallo, R. C. (1985). Genomic diversity of human T-lymphotropic virus type III. *Science* 229, 759-762.
- Zarling, D. A., and Temin, H. M. (1975). High spontaneous mutation rate of an avian sarcoma virus. *J. Virol.* 17, 74-84.