Volume 13 Number 22 1985 Nucleic Acids Research

Polymorphism of the 3' open reading frame of the virus associated with the acquired immune deficiency syndrome, human T-lymphotropic virus type III

Lee Ratner*, Bruno Starcich, Steven F.Josephs, Beatrice H.Hahn, E.Premkumanar Reddy[1], Kenneth J.Livak[2], Stephen R.Petteway,Jr.[2], Mark L.Pearson[2], William A.Haseltine[3], Suresh K.Arya and Flossie Wong-Staal

Laboratory of Tumor Cell Biology, Developmental Therapeutics Program, Division of Cancer Treatment, National Cancer Institute, Bethesda, MD 20205, and [1]Roche Institute, Nutley, NJ 07110, USA

Received 19 August 1985; Revised 10 October 1985; Accepted 11 October 1985

ABSTRACT

The genome of the virus associated with the acquired immune deficiency syndrome (AIDS), human T-lymphotropic virus type III (HTLV-III), includes two open reading frames, not found in other retroviruses. One of these, designated 3'open reading frame (3'orf) is 648 base pairs (bp) in length, and overlaps with the 3'long terminal repeat (LTR) sequences. Sequences of additional HTLV-III clones were determined in order to estimate the level and location of variation within 3'orf, to gain some insight into the function of its protein product. Newly determined sequences are reported for 3'orf of two unintegrated clones of HTLV-III and three cDNA clones made from virion RNA derived from the same cell line infected with pooled blood samples of different patients with AIDS or AIDS-related complex symptoms (ARC). In addition, sequences for 3'orf were derived from an unintegrated viral clone derived from a different cell line infected with a distinct isolate from a single patient. These sequences are compared to those previously reported for six other viral clones. Sequences of 3'orf differ among clones by 1.1-10.4% bp and 2.4-17.0% of predicted amino acids. This represents significantly greater sequence variation than is found in the entire genome on average. Moreover, a functional proviral clone has a termination codon at amino acid residue 124 of this open reading frame. This raises questions concerning the structure, and regulation of expression of the protein encoded by 3'orf.

INTRODUCTION

HTLV-III, also designated lymphadenopathy-associated virus (LAV) or AIDS-related virus (ARV), is closely associated with AIDS and is likely the etiological agent in this disease (1-6). It is similar to other members of the human and bovine lymphotropic virus family with respect to a) tropism for lymphocytes (1,7-13), b) ability to produce giant multinucleated cells in culture (1,11,12,14-16), c) magnesium requirement of its reverse transcriptase (17,18), d) immunological cross-reactivity of some of the gag and env proteins (3,4,19), e) size of its genome which is longer than those of other retroviruses (20-27), and f) ability to give rise to a trans-acting function which enhances viral transcriptional activity (28-33).

HTLV-III differs from HTLV-I and -II in several respects, however. First, HTLV-III has a distinct genetic structure from that of HTLV-I and -II and shares only very distant nucleic acid similarities with them (23,34-36). Second, HTLV-III is strongly cytopathic for T cells, but has no immortalizing activity (1,5,37-39). In contrast, HTLV-I and -II which are

© IRL Press Limited, Oxford, England.

weakly cytopathic (40-42) can efficiently immortalize T lymphocytes in vitro (11,12,43-45). Lastly, HTLV-III shows significant genomic diversity among different isolates unlike HTLV-I (23,36,46,47).

The sequence of the entire provirus of several AIDS viral clones have been presented elsewhere (22-24,27). These data demonstrate that the genome of HTLV-III includes typical retroviral genes for gag, pol, and env, as well as two additional open reading frames not found in other RNA tumor viruses. The first has been designated as the short open reading frame (sor) or as open reading frame 1, Q or P', and overlaps with the 3' portion of pol. The second has been designated 3'orf or open reading frame 2, F, or E'; it begins just 3' to env and overlaps the 3'LTR. Though mRNAs encoding these genes have been identified (22, our unpublished observations), their protein products have not yet been detected. Furthermore, the functions of these putative proteins are unknown.

In order to gain some insight into the mode of expression and function of the 3'orf product, we have determined the sequence of this region from several additional HTLV-III clones and compared them to previously published sequences. Of particular interest is the 3'orf sequence of a functional HTLV-III clone which reveals a termination codon at amino acid position 124. This suggests that the N-terminal portion of this protein product represents the functional portion of the molecule, or that the termination codon is suppressed, perhaps as a method of regulating expression of this viral protein.

MATERIALS AND METHODS

Materials: Nucleotide sequencing reagents, γ-^{32}P-ATP (spec. act. >5000 Ci/mmole) and α-^{32}P-dNTPs (spec. act. >3000 Ci/mmole) were obtained from New England Nuclear; T4 polynucleotide kinase from New England Biolabs; acrylamide, bis-acrylamide, N,N,N',N'-tetramethylethylenediamine, and ammonium persulfate from Biorad; α-^{32}P-dideoxyATP (spec. act. >7000 Ci/mmole) and terminal deoxynucleotidyl transferase from Amersham; restriction enzymes, 10x universal core buffer, urea, bromphenol blue, ethidium bromide, and Klenow fragment of DNA polymerase I from Bethesda Research Labs; and tris (hydroxymethyl)aminomethane and ethylene diamine tetraacetic acid from Sigma.

Phage and Plasmid Clones: Clones HXB2 and HXB3 include sequences from the complete integrated proviruses together with flanking cellular sequences; they were cloned at the Xba I site of λgtwes-λb (36). Clone HAT3 is a 9.2

kilobase (kb) clone derived by Sst I treatment of the Hirt supernatant fraction of the infected cell line; it was cloned in λgtwes-λb (48). C15, C3, and C2 are cDNA clones synthesized from virion RNA using oligo.dT as the primer, which were cytidylate-tailed and cloned in the guanylate-tailed Pst I site of pBR322 (34).

DNA Sequencing: DNA fragments were labeled at their 5'ends with T4 polynucleotide kinase and γ-^{32}P-ATP or at their 3'ends with Klenow fragment of DNA polymerase I and α-^{32}P-dNTP or terminal deoxynucleotidyl transferase and α-^{32}P-dideoxyATP and sequenced by the partial chemical cleavage method (49). The complete 3'orf and LTR sequences of clones HXB2, HXB3, and HAT3 were determined. Clone C15 was also sequenced in its entirety. Sequences determined for clones C3 and C2 were incomplete and consisted of 537 (nucleotides 8302-8838) and 484 bp (nucleotides 8438-8921), respectively.

RESULTS

A permanent T cell line was infected with pooled blood samples from a number of different patients with AIDS or ARC (1). Three integrated proviral clones (HXB2, HXB3, and H9pv22), three unintegrated proviral clones (BH10, BH8, and BH5), and cDNA clones synthesized using oligo.dT as a primer for reverse transcriptase copying of either virion RNA (clones C2, C3, and C15) or cellular polyadenylated RNA (represented as cumulative sequences designated GcDNA) were obtained from this cell line (22,34-36). An unintegrated viral clone (HAT3) was obtained from another T cell line infected with an HTLV-III isolate from a single, different individual (48). Clones LAV1A and ARV2 are also derived from cell lines each infected with distinct viral isolates from individual patients (24,27).

Newly derived sequences for 3'orf and the LTR derived from clones HXB2, HXB3, C15, and HAT3 are shown in Fig. 1. These are compared to previously published sequences of clones BH10, BH8, H9pv22, GcDNA, LAV1A, and ARV2 (22-24,27). The sequences begin with the first ATG codon at position 8374, situated in 3'orf 11 codons after the preceding in-frame termination codon and 1 bp after the termination codon of env. Clone C15 begins at position 8446 and terminates at the polyadenylation site at position 9213. Clones BH10 and BH8 terminate at the Sst I site used for cloning at position 9156.

Clone ARV2 includes a 12 bp insertion in 3'orf at positions 8446-8447, and clone HAT3 includes a 6 bp deletion in 3'orf at positions 8517-8522 relative to HXB2. Neither of these alterations produces a frame shift. Numerous nucleotide differences of the clones compared to HXB2 are distri-

Table 1 - Sequence polymorphisms in 3'orf and the LTR of HTLV-III clones

	Clone	Percent bp differences compared to clone BH10			Percent predicted amino acid differences compared to clone BH10		Percent predicted non-conservative amino acid substitutions compared to clone BH10(a)	
		3'orf	LTR	total(b)	3'orf	total(b)	3'orf	total(b)
Clones derived from the same cell line infected with pooled blood samples of different individuals	BH8/BH5	1.9	1.6	1.3	2.9	1.8	1.9	0.9
	HXB2	2.3	1.6		3.8		1.0	
	HXB3	1.1	1.2		2.4		1.9	
	C2	2.3	2.3		5.6		3.1	
	C3	2.5	2.8		3.9		1.9	
	C15	1.5	1.2		3.3		2.2	
	H9pv22	1.5	1.6	0.5	3.4	0.6	1.4	0.2
Clones each derived from a cell line infected with a viral isolate from a single individual	HAT3	10.4(c)	8.3		17.0(c)		7.3	
	LAV1A	2.1	2.0	1.5	3.9	2.4	1.4	0.8
	ARV2	8.1(d)	5.6	6.3	15.0(d)	10.0	5.3	4.3

a. Excludes deletions and insertions
b. Total nucleotide and predicted amino acid differences determined only for clones or combination of clones e.g. BH8/BH5, for which sequences are available for nearly the entire genome (8937 bp of BH10)
c. Includes 6 bp (2 amino acid) deletion
d. Includes 12 bp (4 amino acid) insertion

buted throughout 3'orf, and do not cluster in any particular region. Many of these sequence polymorphisms are found in common among the different clones.

The level of sequence polymorphisms is summarized in Table 1 for each of the 10 clones compared to BH10, a clone which includes sequences of 98% of the viral genome. Included in the comparison are clones C2 and C3 for which only partial sequences are available, and thus they were not included in Fig. 1 (see Materials and Methods). Nucleotide sequences within 3'orf differ by 1.1-2.5% among the clones isolated from the same cell line infected with pooled blood specimens from different patients. These nucleotide sequence differences predict 2.4-5.6% amino acid differences, of which 1.0-3.1% represent non-conservative amino acid substitutions. Clones HAT3, LAV1A, and ARV2 derived from other individuals show 10.4, 2.1, and 8.1% bp differences, respectively, compared to BH10. There are 17.0, 3.9, and 15.0% predicted amino acid differences in clones HAT3, LAV1A, and ARV2, respectively, compared to BH10; 7.3, 1.4, and 5.3% represent non-conservative amino acid substitutions, respectively. This level of nucleotide and amino acid sequence diversity is significantly greater than that seen on average in the entire viral genome.

Differences in the LTR sequences available for the viral clones are also summarized in Table 1 for comparison to those of 3'orf. This demonstrates 1.2-2.8% nucleotide differences among the clones derived from the same cell line infected from pooled blood from different individuals. Cell

```
HXB2      ATGGGTGGCAAGTGGTCAAAAAGTAGTGTGATTGGATGGCCTACTGTAAGGGAAAGAATGAGACGAGCTGAGC    8446
          MetGlyGlyLysTrpSerLysSerSerValIleGlyTrpProThrValArgGluArgMetArgArgAlaGlu     (24)
HXB3      ------------------------------G-----------G------------------------------
                                        Val         Ala
C15                                                                                -
HAT3      -------------------------AGA---G----------G-----------------CA-AA--------
                                  LysMetGly         Ala               GlnLys
BH10      ------------------------------G-----------G------------------------------
                                        Val         Ala
BH8       ------------------------------G-----------G------------------------------
                                        Val         Ala
H9pv22    ------------------------------------------G------------------------------
                                                    Ala
GcDNA     ------------------------------G-----------G------------------------------
                                        Val         Ala
LAV1A     ------------------------------G------------------------------------------
                                        Val
ARV2      ---------------------C-----A--GG-------T--G--A---------------------------CA
                               Arg   MetGly      SerAlaIle                        Pro

HXB2                CAGCAGCAGATGGGGTGGGAGCAGCATCTCGAGACCTGGAAAAACATGGAGCAATCACAAGTAGC  8511
                   ProAlaAlaAspGlyValGlyAlaAlaSerArgAspLeuGluLysHisGlyAlaIleThrSerSer  (46)
HXB3                -------------------------------------A---------------------------
C15                 ------------------------T------------A---------------------------
                                           Val
HAT3                ----------------------------------------G---------A--------------
                                                                      Thr
BH10                -------------------------------------A---------------------------
BH8                 ------------------------T------------A---------------------------
                                           Val
H9pv22              -------------------------------------A---------------------------
GcDNA               -------------------------------------A---------------------------
LAV1A               -----------------------------------------------------------------
ARV2      CGAGCTGAGC------------------------T----------------------------------------
          ArgAlaGlu                        Val

HXB2      AATACAGCAGCTACCAATGCTGCTTGTGCCTGGCTAGAAGCACAAGAGGAGGAGGAGGTGGGTTTTCCAGTCACA  8586
          AsnThrAlaAlaThrAsnAlaAlaCysAlaTrpLeuGluAlaGlnGluGluGluGluValGlyPheProValThr  (71)
HXB3      --C----------A----------------------------------------A--------------------
                      Asn                                       Lys
C15       --C-----------------C--------T---------------------------------------------
HAT3      -----{    }--AT-----------CA----------------------T----------------------G-
                      Asn            Thr                  Asp                     Arg
BH10      --C----------A--------A----------------------------------------------------
                      Asn      Asp
BH8       --------------------C-A------T---------------------------------------------
                               Asp
H9pv22    --C----------A----------------------------------------A--------------------
                      Asn                                       Lys
GcDNA     ----------------------A----------------------------------------------------
                               Asp
LAV1A     ---------------------------------------------------------------------------
ARV2      --------------T-------A------------------------------A-------------------G-
                               Asp                                                Arg

                                                                                   ./--U3
HXB2      CCTCAGGTACCTTTAAGACCAATGACTTACAAGGCAGCTGTAGATCTTAGCCACTTTTTAAAAGAAAAGGGGGGA  8661
          ProGlnValProLeuArgProMetThrTyrLysAlaAlaValAspLeuSerHisPheLeuLysGluLysGlyGly  (96)
HXB3      ---------------------------------------------------------------------------
C15       -----------------C---------------------------------------------------------
                         Ser
HAT3      -----------------G--T-------T----------------------------------------------
                                     Phe
BH10      ---------------------------------------------------------------------------
BH8       ---------------------------------------------------------------------------
H9pv22    ---------------------------------------------------------------------------
GcDNA     ---------------------------------------------------------------------------
LAV1A     ---------------------------------------------------------------------------
ARV2      ---------------------------------------T-----A-----------------------------
                                                 Leu   Ile

HXB2      CTGGAAGGGCTAATTCACTCCCAAAGAAGACAAGATATCCTTGATCTGTGGATCTACCACACACAAGGCTACTTC  8736
          LeuGluGlyLeuIleHisSerGlnArgArgGlnAspIleLeuAspLeuTrpIleTyrHisThrGlnGlyTyrPhe (121)
HXB3      ------------------------C--------------------------------------------------
C15       ------------------------C--------------------------------------------------
HAT3      -----T------G-GTT------G-A-------------------------G-T---------------------
             Asp      ValPhe      Lys                        Val
BH10      ------------------------C--------------------------------------------------
BH8       ------------------------C-----------------------------C--------------------
                                                                His
H9pv22    ------------------------C--------------------------------------------------
GcDNA     ------------------------C--------------------------------------------------
LAV1A     ------------------------C--------------------------------------------------
ARV2      ---------------TGG-----------------G---------------------------------------
                         Trp               Glu
```

```
HXB2     CCTGATTAGCAGAACTACACACCAGGGCCAGGGGTCAGATATCCACTGACCTTTGGATGGTGCTACAAGCTAGTA  8811
         ProAsp * GlnAsnTyrThrProGlyProGlyValArgTyrProLeuThrPheGlyTrpCysTyrLysLeuVal  (146)
HXB3     -------G------------------A------A-A------------------------C--------------
                Trp                       Ile                        Arg
C15      -------G------------------------A------------------------------------------
                Trp
HAT3     -----C-G-------------------------AC-----------------------------T----------
                Trp                       Thr                            Phe
BH10     ---------------------------------A-----------------------------------------
                                          Ile
BH8      -------G------------------------A------------------------------------------
                Trp
H9pv22   -------G------------------A------A-----------------------------------------
                Trp                       Ile
GcDNA    -------G------------------A------A-----------------------------------------
                Trp                       Ile
LAV1A    -------G-------------------------------------------------------------------
                Trp
ARV2     -------G------T------------------A------------------------------T----------
                Trp                       Ile                            Phe

HXB2     CCAGTTGAGCCAGATAAGATAGAAGAGGCCAATAAAGGAGAGAACACCAGCTTGTTACACCCTGTGAGCCTGCAT  8886
         ProValGluProAspLysIleGluGluAlaAsnLysGlyGluAsnThrSerLeuLeuHisProValSerLeuHis  (171)
HXB3     --------------G---T-------A-----C------------------------------------------
                     Glu   Leu
C15      ------------------G-------------C------------------------------------------
                           Val
HAT3     ------------------G------------C-G------------A----------------A-AT--------
                           Val         ThrGlu         Asn               IleCys
BH10     --------------G---T-------A-----C------------------------------------------
                     Glu   Leu
BH8      --------------G-----------A------------------------------------------------
                     Glu
H9pv22   ------------------G-------------C------------------------------------------
                           Val
GcDNA    --------------G---T-------A-----C------------------------------------------
                     Glu   Leu
LAV1A    ------------------G--------------------------------------------------------
                           Val
ARV2     --------------G---G--------------G------------A----------------A-----------
                     Glu   Val            Glu         Asn               Met

HXB2     GGGATGGATGACCCGGAGAGAGAAGTGTTAGAGTGGAGGTTTGACAGCCGCCTAGCATTTCATCACGTGGCCCGA  8961
         GlyMetAspAspProGluArgGluValLeuGluTrpArgPheAspSerArgLeuAlaPheHisHisValAlaArg  (196)
HXB3     --A------------------------------------------------------------------------
C15      --A---------------------------------------------------------------A--------
                                                                           Met
HAT3     --------------A----A-----------T-----A---------------C--------------C------
                           Lys         Val   Lys
BH10     --A---------------------------------------------------------------A--------
                                                                           Met
BH8      --A-----------T---------------------------------------------------A--------
                                                                           Met
H9pv22   --A------------------------------------------------------------------------
GcDNA    --A------------------------------------------------------------------------
LAV1A    --A-----------T------------------------------------------------------------
ARV2     --------G---G------A-----------T----------------AAA---------------A--------
               Glu   Ala   Lys         Val               Lys               Met

HXB2     GAGCTGCATCCGGAGTACTTCAAGAACTGCTGACATCGAGCTTGCTACAAGGGACTTTCCGCTGGGGACTTTCCA  9036
         GluLeuHisProGluTyrPheLysAsnCys                                               (206)
HXB3     ---------------------------------T-----------------------------------------
C15      ---------------------------------------------------------------------------
HAT3     ---AA--------------A---AG---------------T--T-------------------------------
            Lys            Tyr   Asp
BH10     ---------------------------------------------------------------------------
BH8      ---------------------------------T-----------------------------------------
H9pv22   ---------------------------------T-----------------------------------------
GcDNA    ---------------------------------T-----------------------------------------
LAV1A    ---------------------------------------------------------------------------
ARV2     -------------------A---AG------------------T-------------------------------
                           Tyr   Asp

HXB2     GGGAGGCGTGGCCTGGGCGGGACTGGGGAGTGGCGAGCCCTCAGATCCTGCATATAAGCAGCTGCTTTTTGCCTG  9111
HXB3     ---------------------------------------------------------------------------
C15      ---------------------------------------------------------------------------
HAT3     ------T---------------------------------------G----------------------------
BH10     ---------------------------------------------------------------------------
BH8      ---------------------------------------------------------------------------
H9pv22   ----------------------------------------------G----------------------------
GcDNA    ---------------------------------------------------------------------------
LAV1A    ----------------------------------------------G----------------------------
ARV2     ------------------------------------T---------G----------------------------
```

```
         U3--\/--R          .         .         .         .         .         .
HXB2      TACTGGGTCTCTCTGGTTAGACCAGATCTGAGCCTGGGAGCTCTCTGGCTAACTAAGGAACCCACTGCTTAAGCC  9186
HXB3      ---------------------------------------------------------------------------
C15       ---------------------------------------------------------------------------
HAT3      ---------------------------------T-----------------G-----------------------
BH10      -------------------------------------------
BH8       -------------------------------------------
H9pv22    -------------------------------------------------------GA------------------
GcDNA     -------------------------------------------------------GA------------------
LAV1A     ---------------------------T---------------------------G-------------------
ARV2      -------------------------------------------------------G-------------------
             .         .         R--\/--U5 .         .         .         .         .
HXB2      TCAATAAAGCTTGCCTTGAGTGCTTCAAGTAGTGTGTGCCCGTCTGTTGTGTGACTCTGGTAACTAGAGATCCCT
HXB3      ---------------------------------------------------------------------------
C15       ----------------------------AAAAAA
HAT3      ---------------------------------------------------------------------------
H9pv22    ---------------------------------------------------------------------------
GcDNA     -----------------------------------------A---------------------------------
LAV1A     ---------------------------------------------------------------------------
ARV2      ---------------------------------------------------------------------------
                  .         .        .   U5--\
HXB2      CAGACCCTTTTAGTCAGTGTGGAAAA TCTCTAGCA
HXB3      -------------------------- ---------
HAT3      ------TC------------------ ---------
H9pv22    -------------------------- ---------
GcDNA     -------------------------- ---------

LAV1A     -------------------------- ---------
ARV2      --------------------------A---------
```

Fig. 1. 3'Open reading frame and long terminal repeat sequences of HTLV-III clones (listed on the left of each line). Newly determined sequences for clones HXB2, HXB3, C15, and HAT3 (listed at the top of each group of lines) are compared to previously published sequences for clones BH10, BH8, H9pv22, GcDNA, LAV1A, and ARV2 (listed at the bottom of each group of lines). GcDNA represents cumulative sequences from several different cDNA clones made from cellular polyadenylated RNA. Nucleotide and predicted amino acid sequences are shown for clone HXB2, and dashes represent identical nucleotides at the same position in the other clones. Sequence differences in the other clones are indicated. LTR sequences for HXB2 and HXB3 were determined for both the 5' and 3' LTR of the integrated proviruses and are identical in each case. The sequences shown for U5 and a portion of R of HAT3 were determined from the 5'LTR and are inferred to be the same for the 3'LTR. The sequence of the 5'LTR of HXB2 was previously reported (53). Sequences shown begin with the first ATG codon in 3'orf, 11 codons after an in-frame termination codon. Amino acid sequence positions for 3'orf are listed to the right of each group of lines in parentheses. Nucleotide positions are also listed to the right of each group lines, using the previously described numbering scheme where position 1 represents the viral RNA CAP site in the 5'LTR and position 9213 represents the polyadenylation site in the 3'LTR (23). Dots above each group of sequences represents the position of every tenth nucleotide in clone HXB2. An asterisk indicates a termination codon at amino acid position 124 in 3'orf of clone HXB2. A deletion in clone HAT3 relative to HXB2 is indicated by brackets. A 1 bp deletion in clone ARV2 relative to HXB2 at position 9071 is indicated by a blank space. The boundaries for U3, R, and U5 are also depicted.

lines infected with isolates from other individuals show 8.3, 2.0, and 5.6% differences in LTR sequences for clones HAT3, LAV1A, and ARV2, respectively, compared to BH10.

Of note is that HXB2, which has been demonstrated to be a functional clone with regards to its ability to give rise to cytopathic viral particles (50), has a TAG termination codon at amino acid residue 124 (nucleotide positions 8743-8745) of 3'orf (Fig. 1). A TAG termination codon is found at the same position in clones BH10 (Fig. 1) and C2 and C3 (data not shown). The other clones have a TGG sequence coding for tryptophan at that posi-

tion. Thus, at most a 14 kilodalton primary translation product could be encoded by clone HXB2 if this termination codon is not suppressed.

DISCUSSION

Sequences compared here include those from nine different HTLV-III viral clones derived from a single cell line infected with pooled blood samples from different individuals. Polymorphism in sequences is clearly present with 1.2-2.8% variation within LTR sequences among these clones. Other clones derived from different cell lines infected with distinct single viral isolates reveal 2.0-8.3% bp differences in the LTR compared to clone BH10. The contribution of cloning artifacts or sequencing errors to this level of variation is less than 0.1% bp, as was previously demonstrated (24,47).

These sequence polymorphisms confirm an earlier report of genomic diversity of HTLV-III based on restriction enzyme site differences among different isolates (36). In that study, differences in restriction enzyme maps were noted in the fresh tissues of different individuals as well as in cell lines. There is no evidence for variation in genomic sequences as a result of in vitro propagation of virus infected cells, though definitive data are still lacking. Sequence variations may represent differences in sequences of viruses infecting different individuals or differences arising within a given individual over a period of time. In one case examined, however, an identical restriction enzyme map was found for virus clones isolated from the same individual on two different occasions several months apart (51). The mechanisms involved in the generation of sequence diversity remain to be investigated. An error-prone reverse transcriptase may account for these results. Moreover, the implications for differences in function of viral proteins is not known. The resultant alteration in antigenic structure of viral proteins may be important in determining the interaction of infected cells with immune regulators, and will also influence strategies of vaccine development.

An open reading frame of 648 bp, designated 3'orf, is present in some but not all AIDS viral clones. The first ATG codon follows the env gene, and this open reading frame overlaps with U3 by 331 bp. Clones from the same cell line infected with pooled blood samples show 1.1-2.5% bp differences and 2.4-5.6% predicted amino acid differences compared to BH10. Clones HAT3, LAV1A, and ARV2, derived from other cell lines infected with isolates from different single individuals show 10.4, 2.1, and 8.1% bp

differences, and 17.0, 3.9, and 15.0% predicted amino acid differences, respectively, compared to BH10. This level of sequence diversity is greater than that seen on average in the entire viral genome. For example, in comparing ARV2 with BH10, the level of predicted amino acid differences are 6.4% for gag, 4.8% for pol, 7.9% for sor, 17.0% for the extracellular portion of env, 12.2% for the transmembrane portion of env, and 15.0% for 3'orf (23,24,47). The high level of sequence diversity in 3'orf compared to most of the other regions of the provirus, however, does not exclude the possibility that a protein product from this region may be functionally important. Certainly the env product, which also exhibits considerable sequence variability, is required for virus replication. However, it does raise the question as to whether a 3'orf product may contribute to antigenic variation of the virus.

Several clones were also noted to have a termination codon at the predicted amino acid residue 124 in 3'orf. Moreover, clone HXB2, one of the clones with a termination codon at this position, upon transfection into human umbilical cord blood cells gives rise to viral particles and cytopathic effects, and thus appears to be a fully functional clone (50). These data raise questions about the structure of a protein encoded by 3'orf and its possible biological functions. The N terminal portion of the 3'orf protein product may include the functional region of the molecule. The possibility that the termination codon at residue 124 in some viral isolates is suppressed must also be considered. The demonstration of suppression of the termination codon between gag and pol of Moloney murine leukemia virus provides a precedent among retroviruses for this mechanism (52). Thus, suppression of termination codons may provide an additional level of regulation of viral protein expression.

ACKNOWLEDGEMENTS

We thank E. Whitehorne, A. Rafalski, G. Chan, and G. Shaw for assistance with these experiments and helpful discussions, and we thank R.C. Gallo for his continuing support. Part of this work was suppoted by NIH grant CA36974 and ACS grant RD-186.

*To whom reprint requests should be addressed. Present address: Division of Hematology and Oncology, Washington University, Box 8125, 660 S. Euclid, St. Louis, MO 63110, USA

[2]Central Research and Development Department, Experimental Station, E.I. DuPont de Nemours, Wilmington, DE 19898, and [3]Dana Farber Cancer Institute and Department of Pathology, Harvard Medical School, Department of Cancer Biology, Harvard School of Public Health, Boston, MA 02114, USA

REFERENCES

1. Popovic, M., Sarngadharan, M.G., Read, E., and Gallo, R.C. (1984) Science 224, 497-500.
2. 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) Science 224, 500-503.
3. Sarngadharan, M.G., Popovic, M., Bruch, L, Schupbach, J., and Gallo, R.C. (1984) Science 224, 506-508.
4. Schupbach, J., Popovic, M., Gilden, R.V., Gonda, M.A., Sarngadharan, M.G., and Gallo, R.C. (1984) Science 224, 503-505.
5. Barre-Sinoussi, F., Chermann, J-C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., Dauguet, C., Axler-Blin, C., Vezinet-Brun, F., Rouzioux, C., Rozenbaum, W., and Montagnier, L, (1983) Science 220, 868-870.
6. Levy, J.A., Hoffman, A.D., Kramer, S.M., Landis, J.A., Shimbukuro, J.M., and Oshiro, L.S. (1984) Science 225, 840-842.
7. Burny, A., Bruck, C., Chantrenne, H., Cleuter, Y., Dekezel, D., Ghysdael, J., Kettmann, R., Leclercq, M., Leunen, J., Mammerickx, M., and Portelle, D. (1980) in Viral Oncology, Klein, G. Ed., Vol. 1, pp. 231-280, Raven Press, New York.
8. Gallo, R.C., Mann, D., Broder, S., Ruscetti, F.W., Maeda, M., Kalyanaraman, V.S., Robert-Guroff, M., and Reitz, M.S. (1982) Proc. Natl. Acad. Sci. USA 79, 5680-5683.
9. Chen, I.S.Y., Quan, S.G., and Golde, D.W. (1983) Proc. Natl. Acad. Sci. USA 80, 7006-7009.
10. Dalgleish, A.G., Beverley, P.C.L., Clapham, P.R., Crawford, D.H., Greaves, M.F., and Weiss, R.A. (1984) Nature 312, 763-767.
11. Popovic, M., Lange-Wantzin, G., Sarin, P.S., Mann, D., and Gallo, R.C. (1983) Proc. Natl. Acad. Sci. USA 80, 5402-5406.
12. Popovic, M., Sarin, P.S., Robert-Guroff, M., Kalyanaraman, V.S., Mann, D., Minowada, J., and Gallo, R.C. (1983) Science 219, 856-859.
13. Klatzmann, D., Barre-Sinoussi, F., Nugeyre, M.T., Dauget, C., Vilmer, E., Griscelli, C., Brun-Vezinet, F., Rouzioux, C., Gluckman, J.C., Chermann, J-C., and Montagnier, L. (1984) Science 225, 59-63.
14. Hoshino, H., Esumi, H., Miwa, M., Shimoyama, M., Minato, K., Tobinai, K., Hirose, M., Watanabe, S., Inada, N., Kinoshita, K., Kamihira, S., Ichimaru, M., and Sugimura, T. (1983) Proc. Natl. Acad. Sci. USA 80, 6061-6065.
15. Nagy, K., Clapham, P., Cheingsong-Popov, R., and Weiss, R.A. (1983) Int. J. Cancer 32, 321-328.
16. Clapham, P., Nagy, K., Cheingsong-Popov, R., and Weiss, R.A. (1984) Proc. Natl. Acad. Sci. USA 81, 2886-2889.
17. Rho, H.M., Poiesz, B., Ruscetti, F.W., and Gallo, R.C. (1981) Virology 112, 355-360.
18. Rey, M.A., Spire, B., Dormont, D., Barre-Sinoussi, F., Montagnier, L, and Chermann, J.C. (1984) Biochem. Biophys. Res. Commun. 121, 126-133.
19. Sarngadharan, M.G., Bruch, L., Popovic, M., and Gallo, R.C. (1985) Proc. Natl. Acad. Sci. USA 82, 3481-3484.
20. Seiki, M. Hattori, S., Hirayama, V., and Yoshida, M. (1983) Proc. Natl. Acad. Sci. USA 79, 6899-6902.
21. Shimotohno, K., Wachsman, W., Takahashi, Y., Golde, D.W., Miwa, M., Sugimura, T., and Chen, I.S.Y. (1984) Proc. Natl. Acad. Sci. USA 81, 6657-6661.
22. Muesing, M.A., Smith, D.H., Cabradilla, C.D., Benton, C.V., Lasky, L.A., and Capon, D.J. (1985) Nature 313, 450-458.
23. Ratner, L., Haseltine, W., Patarca, R., Livak, K.J., Starcich, B., Josephs, S.F., Doran, E.R., Rafalski, J.A., Whitehorn, E.A., Baumeister, K., Ivanoff, L., Petteway, S.R., Jr., Pearson, M.L., Lautenberger, J.A., Papas, T.S., Ghrayeb, J., Chang, N.T., Gallo, R.C., and Wong-Staal, F. (1985) Nature 313, 277-284.
24. Sanchez-Pescador, R., Power, M.D., Barr, P.J., Steimer, K.S., Stempien, M.M., Brown-Shimer, S.L., Gee, W.W., Renard, A., Randolph, A., Levy, J.A., Dina, D., and Luciw, P.A. (1985) Science 227, 484-492.
25. Sagata, N., Yasunaga, T., Tsuzuku-Kawamura, J., Ohishi, K., Ogawa, Y., and Ikawa, Y. (1985) Proc. Natl. ACad. Sci. USA 81, 4741-4745.
26. Shimotohno, K., Takahashi, Y., Shimzu, N., Gojobori, T., Golde, D.W., Chen, I.S.Y., Miwa, M., and Sugimura, T. (1985) Proc. Natl. Acad. Sci. USA 82, 3101-3105
27. Wain-Hobson, S., Sonigo, P., Danios, O., Cole, S., and Alizon, M. (1985) Cell 40, 9-17.
28. Sodroski, J.G., Rosen, C.A., and Haseltine, W.A. (1984) Science 225, 381-385.
29. Arya, S.K., Guo, C., Josephs, A.F., and Wong-Staal, F. (1985) Science 229, 69-73.
30. Rosen, C.A., Sodroski, J., Kettman, R., Burny, A., and Haseltine, W.A. (1985) Science 227, 320-321.

31. Rosen, C., Sodroski, J.G., and Haseltine, W.A. (1985) Nature, in press.
32. Sodroski, J.G., Rosen, C., Wong-Staal, F., Salahuddin, S.Z., Popovic, M., Arya, S., Gallo, R.C., and Haseltine, W.A. (1985) Science 227, 171-173.
33. Sodroski, J., Patarca, R., Rosen, C., Wong-Staal, F., and Haseltine, W. (1985) Science 229, 74-77.
34. Arya, S.K., Gallo, R.C., Hahn, B., Shaw, G., Popovic, M., Salahuddin, S.Z., and Wong-Staal, F. (1984) Science 225, 927-930.
35. Hahn, B.H., Shaw, G.M., Arya, S.K., Popovic, M., Gallo, R.C., and Wong-Staal, F. (1984) Nature 312, 166-169.
36. Shaw, G.M., Hahn, B.H., Arya, S.K. Groopman, J.E., Gallo, R.C., and Wong-Staal, F. (1984) Science 226, 1165-1171.
37. Gottlieb, M.S., Schroff, R., Schanker, H.M, Weisman, J.D., Fan, P.T., Wolf, R.A., and Saxon, A. (1981) New Engl. J. Med. 305, 1425-1431.
38. Laurence, J., Brun-Vezinet, F., Schutzer, S.E., Rouzioux, C., Klatzmann, D., Barre-Sinoussi, F., Chermann, J-C., and Montagnier, L. (1984) New Engl. J. Med. 311, 1269-1273.
39. Safai, B., Sarngadharan, M.G., Groopman, J.E., Arnett, K., Popovic, M., Sliski, A., Schupbach, J., and Gallo, R.C. (1984) Lancet i, 1438-1440.
40. Essex, M.E., McLane, M.F., Tachibana, N., Francis, D.P., and Lee, T-H. (1984) in Human T-Cell Leukemia-Lymphoma Virus, Gallo, R.C., Essex, M. and Gross, L., Eds. pp. 355-362, Cold Spring Harbor Laboratory, New York.
41. Popovic, M., Flomenberg, N., Volkman, D.J., Mann, D., Fauci, A.S., DuPont, B., and Gallo, R.C. (1984) Science 226, 459-461.
42. DeVecchis, L., Graziani, G., Macchi, B., Grandori, C., Pastore, S., Popovic, M., Gallo, R.C., and Bonmassar, E. (1985) Leuk. Res., in press.
43. Miyoshi, I., Kubonishi, I., Yoshimoto, S., Akagi, T., Ohtsuki, Y., Shiraishi, Y., Nagato, K., and Hinuma, Y. (1981) Nature 294, 770-771.
44. Yamamoto, N., Okada, M., Koyanagi, Y., Kanagi, M., and Hinuma, Y. (1982) Science 217, 737-739.
45. Markham, P.D., Salahuddin, S.Z., Kalyanaraman, V.S., Popovic, M., Sarin, P.S., and Gallo, R.C. (1983) Int. J. Cancer 31, 413-420.
46. Gallo, R.C. (1984) Cancer Surv. 3, 113-159.
47. Ratner, L., Gallo, R.C., and Wong-Staal, F. (1985) Nature 313, 636-637.
48. Hahn, B., Gonda, M.A., Shaw, G.M., and Wong-Staal, F. (1985) Proc. Natl. Acad. Sci. USA, 82, 4813-4817.
49. Maxam, A. and Gilbert, W. (1980) Meth. Enzymol. 65, 499-560.
50. Fisher, A.G., Collalti, E., Ratner, L., Gallo, R.C., and Wong-Staal, F. (1985) Nature 316, 262-265.
51. Wong-Staal, F, Shaw, G.M., Hahn, B.H., and Gallo, R.C. (1985) Science, 229, 759-762.
52. Yoshinaka, Y., Katoh, I., Copeland, T.D., and Oroszolan, S. (1985) Proc. Natl. Acad. Sci. USA 82, 1618-1622.
53. Starcich, B., Ratner, L., Josephs, S.F., Okamoto, T., Gallo, R.C., and Wong-Staal, F. (1985) Science, 227, 538-540.