*Proc. Natl. Acad. Sci. USA*
Vol. 81, pp. 6657–6661, November 1984
Biochemistry

# Nucleotide sequence of the 3′ region of an infectious human T-cell leukemia virus type II genome

**(retrovirus/DNA sequence/conserved amino acid sequence)**

KUNITADA SHIMOTOHNO*, WILLIAM WACHSMAN†, YURI TAKAHASHI*, DAVID W. GOLDE†, MASANAO MIWA*, TAKASHI SUGIMURA*, AND IRVIN S. Y. CHEN†

*National Cancer Center Research Institute, Chuo-ku, Tokyo 104, Japan; and †Division of Hematology and Oncology, Department of Medicine, University of California School of Medicine, Los Angeles, CA 90024

*Contributed by Takashi Sugimura, July 12, 1984*

ABSTRACT **The nucleic acid sequence of the 3′ region of human T-cell leukemia virus type II (HTLV-II) proviral DNA was determined using a HTLV-II proviral clone that could be recovered as infectious, transforming virus. The sequence data indicate a region of unknown function of ≈1.6 kilobase pairs in the 3′ region, analogous to the X region previously identified in human T-cell leukemia virus type I (HTLV-I). Three overlapping open reading frames are present in the X region of HTLV-II. One of these open reading frames, Xc, is most likely to encode a protein product, because it has greater predicted amino acid sequence homology (78%) with the X-IV region of HTLV-I and a greater percentage of its base differences with X-IV at the third nucleotide position of codons than do the other open reading frames. Sequences of the X-region that include the open reading frames are conserved in two deletion mutants of HTLV-II, which are associated with a subline of Mo cells with a decreased dependence on fetal bovine serum.**

Human T-cell leukemia viruses (HTLV) are associated with certain forms of human leukemias and lymphomas (1–5). At least two types of HTLV have been identified. HTLV type I (HTLV-I) is endemic to various regions of the world and is often associated with aggressive leukemias/lymphomas of mature T lymphocytes (3–5). HTLV type II (HTLV-II) was found in a single patient (Mo) with a T-cell variant of hairy-cell leukemia (1, 2, 6). This patient is alive and well 8 yr after splenectomy.

Both HTLV-I and HTLV-II transform normal human peripheral blood or cord blood T lymphocytes *in vitro* (7–10). These virus-transformed T cells have a helper-inducer phenotype similar to that of leukemic cells in patients with HTLV-associated disease. Elucidation of the mechanism of *in vitro* transformation is relevant to the process of leukemogenesis. However, the regions of the HTLV genome necessary for transformation have not been identified. Nucleic acid sequence analysis of the complete HTLV-I genome revealed a region at the 3′ locus of the genome with no known function and without precedent in animal retroviruses other than bovine leukemia virus (11). This region, referred to as X, is suspected to encode protein(s) involved in the process of transformation. The X region does not cross-hybridize with normal human cellular DNA sequences and, therefore, does not encode a retroviral oncogene.

HTLV-II has *in vitro* biological properties similar to but only limited homology with HTLV-I as determined by hybridization of the genomes and nucleic acid sequencing of the long terminal repeat (LTR) (12, 13). By nucleic acid sequence analysis we have identified a region comparable to X in an infectious and transformation-competent molecular clone of HTLV-II. The homology between the two viruses in this region was determined.

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

## MATERIALS AND METHODS

**Sequencing of HTLV-II DNA.** Bacterial plasmid pH6-B3.5, which contains *env*, X, and a part of the LTR of HTLV-II, was used as a source of DNA. pH6-B3.5 was subcloned from a cloned infectious HTLV-II provirus, λH6. The sequencing method of Maxam and Gilbert was applied to 5′- or 3′-end-labeled DNA fragments obtained by digestion of the DNA with restriction enzymes (14). Both strands of the DNA were sequenced.

**Comparison of Nucleotide and Amino Acid Sequences.** Nucleotide or amino acid sequence homology was assessed using a computer program developed by Japan Soft Development.

**Transfection of HTLV-II Proviral DNA.** The procedures used for transfection of lymphoid cells were described previously (15).

**Materials.** Restriction enzymes were purchased from Takara Shuzo (Kyoto, Japan), Bethesda Research Laboratories, or New England Biolabs. Polynucleotide 5′-hydroxyl-kinase and the large fragment of DNA polymerase I were from Boehringer Mannheim and Takara Shuzo, respectively. Radiolabeled nucleotides were from Amersham.

## RESULTS

**Nucleic Acid Sequence Analysis of the 3′ Region of the HTLV-II Genome.** Previous nucleic acid sequence analysis of HTLV-I revealed four potential open reading frames beginning with methionine codons in the 3′ region of *env* (11). However, as the sequenced provirus was not recovered as an infectious virus it may not represent the genome of a replication- and transformation-competent HTLV-I. Therefore, we first determined whether an apparently complete HTLV-II provirus clone, λH6, could be recovered as infectious virus capable of transforming normal human T lymphocytes.

Since HTLV-II can replicate in some B-lymphoblastoid cell lines, we used a B-cell line for HTLV-II DNA transfection. The HTLV-II provirus was subcloned into the plasmid vector pSV2-neo. Protoplasts of *E. coli* HB101 containing the HTLV-II subclone, pH6-neo (Fig. 1*A*), were fused with WIL-2 cells and antibiotic G418-resistant clones of cells were subsequently selected. Of these G418-resistant B-cell clones, ≈25% expressed viral p19 and p24 antigens, as determined by indirect immunofluorescence (Fig. 1*B*), and viral RNA which was correctly initiated from the cap site of the LTR (15). These B cells were lethally irradiated and cocultivated with normal human peripheral blood lymphocytes as

Abbreviations: HTLV-I and HTLV-II, human T-cell leukemia virus types I and II; LTR, long terminal repeat; kbp, kilobase pair(s).

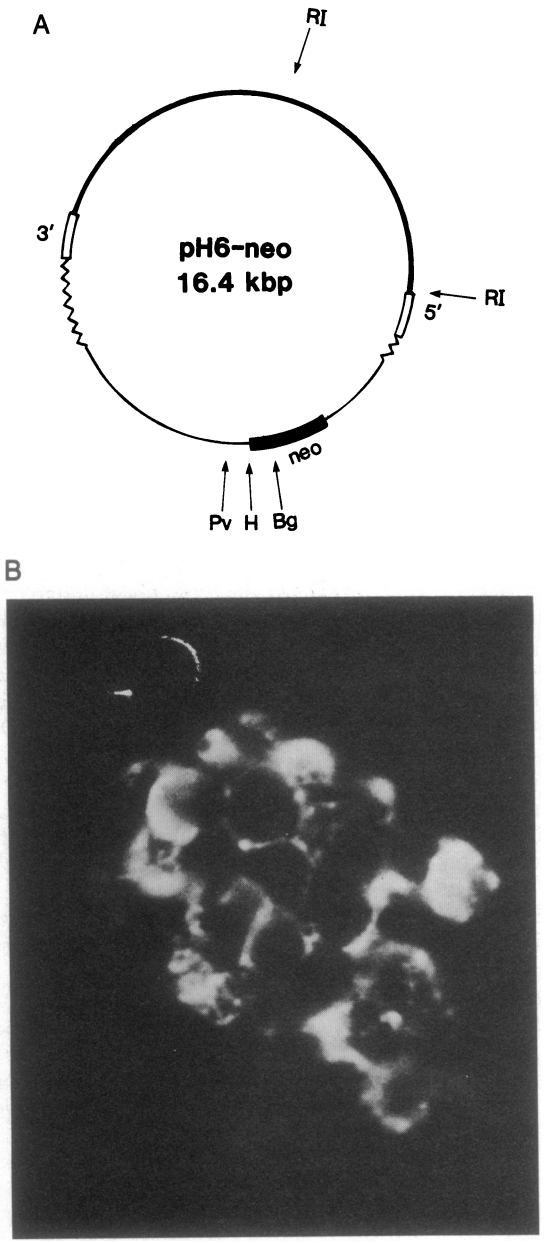


FIG. 1. Transfection and expression of HTLV-II proviral DNA in human B cells. (*A*) The plasmid pH6-neo used for transfection is shown schematically. The complete HTLV-II provirus of λH6 (12) between flanking *Hin*dIII sites in cellular DNA was subcloned into the *Eco*RI site of the plasmid vector pSV2-neo. The DNAs cleaved with *Hin*dIII or *Eco*RI were treated with nuclease S1 to blunt the ends and were ligated. WIL-2 cells were transfected with pH6-neo by spheroplast fusion (15) and stable transformants were propagated. The thick line and open boxes represent the provirus. Zig-zag lines show cellular flanking sequences. The thin lines and shaded box represent pSV-neo sequences. *Bgl* II (Bg), *Hin*dIII (H), *Eco*RI (RI), and *Pvu* II (Pv) sites are indicated. (*B*) One of the stable B-cell transformants that express HTLV-II p24 antigens is shown. Fixed cells were treated with rabbit antisera directed against viral p24 antigens and visualized by indirect immunofluorescence.

described to test for HTLV-II transformation (10). Cell lines with the T-helper surface-antigen phenotype were established and shown to be infected with HTLV-II. No B-cell markers (surface membrane immunoglobulin, Epstein–Barr-virus nuclear or capsid antigen) were detected in the transformed peripheral blood cells. The nucleic acid sequence of the λH6 provirus therefore represents the genome of a replication- and transformation-competent HTLV-II.

The nucleotide sequence of the 3′ region of HTLV-II is shown in Fig. 2; *env* ends one base before the first nucleotide of Fig. 2 and the X region of HTLV-II extends to position 1559.

**Sequence Homology Between the X Regions of HTLV-II and HTLV-I.** There is considerable nucleotide sequence homology between the X regions of HTLV-II and HTLV-I. Most of this sequence homology is in the 3′ two-thirds of the X region that is coincident with the position of the major open reading frames in the X region (see below). In this region there is 75% nucleic acid sequence homology. In contrast there is only 33% sequence homology in the 5′ one-third of the X region. The 5′ region corresponds to the X-I region of HTLV-I (11). No long open reading frames are present in this region of HTLV-II.

Three open reading frames with overlapping sequences were identified in the nucleotide sequence of HTLV-II. These open reading frames Xa, Xb, and Xc correspond to the open reading frames X-II, X-III, and X-IV of HTLV-I (11). The derived amino acid sequence homologies for the codons between flanking termination codons are 62, 61, and 78% for Xa/X-II, Xb/X-III, and Xc/X-IV, respectively. Xc and X-IV share a stretch of 335 corresponding amino acid codons uninterrupted by termination codons (Fig. 3), compared with 96 for Xa/X-II and 145 for Xb/X-III. Furthermore, there is 82% amino acid homology in the 112 codons upstream of the first conserved methionine codon of Xc/X-IV, indicating that the Xc region may encode a fused protein whose initiation codon is in a different region of the genome.

**Frequency of Base Changes at the Third Nucleotide Position in the Open Reading Frames.** The frequency of base changes between divergent sequences in each of the three positions in amino acid codons has been used as a measure of evolutionary conservation and, therefore, functional significance of an open reading frame: the greater the frequency of third nucleotide changes relative to the first and second, the more likely that the reading frame encodes a protein that has been conserved during evolution.

The frequency of base changes in the open reading frames of HTLV was calculated by comparing the sequences of the corresponding open reading frames in HTLV-I and HTLV-II. The number of mismatched bases between the Xa region (from nucleotides 530 to 817) in HTLV-II and the corresponding region in the open reading frame for X-II in HTLV-I is 58 mismatched nucleotides out of 288 nucleotides (only amino acid codons between termination codons are included in this calculation). Fourteen percent of these mismatches are at the third position of codons in this reading frame. In the Xb/X-III region, 28% of the mismatched bases are at the third position. However, for Xc/X-IV, 66% of the mismatched bases are located at the third position, a significantly greater frequency than that of either of the other two reading frames. Thus, Xc is most likely to encode a protein product.

**Conservation of the 3′ Region of HTLV-II DNA in Deletion Mutants.** Molecular cloning and characterization of HTLV-II DNA from Mo cells demonstrated the presence of three forms of HTLV-II proviruses in these cells. The largest cloned provirus represents the complete replication-competent genome of HTLV-II as evidenced by the recovery of infectious, transforming HTLV-II by DNA transfection. The other two forms of HTLV-II DNA were defective, having large internal deletions of the viral genome. However, their LTRs were intact and both defective genomes could be packaged as infectious virus (12). The defective viruses are associated only with a subline of the HTLV-II-infected Mo cells that has growth properties distinct from the original Mo cell line; these cells have a decreased dependence on fetal bovine serum and clone spontaneously in methylcellulose and by limiting dilution.

Restriction enzyme analysis showed that the larger defec-


FIG. 2. Nucleotide sequence homology between the X regions of HTLV-II and HTLV-I proviruses. The nucleotide next to the 3′ end of the *env* gene is designated nucleotide 1. The open reading frames Xa, Xb, and Xc in HTLV-II and X-I, X-II, X-III, and X-IV in HTLV-I are shown. The 5′ portion of the X-I open reading frame is in *env*. Xc and X-IV end in the LTRs. A putative splice acceptor site is indicated by an arrow. Asterisks indicate nucleotides that are identical in the two sequences.

tive genome, typified by clone H9 (Fig. 4), has a deletion of ≈2.0 kilobase pairs (kbp) with conservation of ≈5.0 kbp in the 5′ region and ≈2.0 kbp in the 3′ region. The smaller defective provirus, typified by clone H2, has a deletion of nearly the entire internal sequence of HTLV-II. Excluding the LTR, <2.0 kbp of the sequence is conserved. Detailed restriction enzyme analysis demonstrated that most of the conserved sequence in the 3′ region is the X region of HTLV-II.

```
HTLV-II : LQSCLLSAHFLG FGQSLLYGYP VYVFGDCVQA DWCPVSGGLC STRLHRHALL ATCPEHQLTW   62
          ******* * ****** *** *********  **** ***** * ******** ******* **
HTLV-I  : --PCLLSAHFPG FGQSLLFGYP VYVFGDCVQG DWCPISGGLC SARLHRHALL ATCPEHQITW   60

HTLV-II :   DPIDGRVVSS PLQYLIPRLP SFPTQRTSRT LKVLTPPTTP VSPKVPPAFF QSMRKHTPYR  122
            *******  *  ** ****** ******** * ******* *    *  ** *  * ***  * *
HTLV-I  :   DPIDGRVIGS ALQFLIPRLP SFPTQRTSKT LKVLTPPITH TTPNIPPSFL QAMRKYSPFR  120

HTLV-II :   NGCLEPTLGD QLPSLAFPEP GLRPQNIYTT WGKTVVCLYL YQLSPPMTWP LIPHVIFCHP  182
            **  *****   ** * ** * ****** **  **  *** ** ****** *** * ********
HTLV-I  :   NGYMEPTLGQ HLPTLSFPDP GLRPQNLYTL WGGSVVCMYL YQLSPPITWP LLPHVIFCHP  180

HTLV-II :   RQLGAFLTKV PLKRLEELLY KMFLHTGTVI VLPEDDLPTT MFQPVRAPCI QTAWCTGLLP  242
            ******* * * ** ***** *  * **  *  **** ****  *** ***    ***  ****
HTLV-I  :   GQLGAFLTNV PYKRIEELLY KISLTTGALI ILPEDCLPTT LFQPARAPVT LTAWQNGLLP  240

HTLV-II :   YHSILTTPGL IWTFNDGSPM ISGPYPKAGQ PSLVVQSSLL IFEKFETKAF HPSYLLSHQL  302
            ** ****** **** ** ** **** ** ** **** ***   ** ** ***  *** **** *
HTLV-I  :   FHSTLTTPGL IWTFTDGTPM ISGPCPKDGQ PSLVLQSSSF IFHKFQTKAY HPSFLLSHGL  300

HTLV-II :   IQYSSFHNLH LLFDEYTNIP VSILFNKEEA DDNGD                            337
            ******* ** *** ******  * ***  ** ***
HTLV-I  :   IQYSSFHSLH LLFEEYTNIP ISLLFNEKEA DDNDHEPQIS PGGLEPPSEK HFRETEV     357
```

FIG. 3. Homology of predicted amino acid sequences encoded by the open reading frames Xc (HTLV-II) and X-IV (HTLV-I). Asterisks indicate identical amino acids in the two sequences. Amino acids are represented by standard one-letter abbreviations (16).

Furthermore, the deletion endpoints occur at the 5′ end of the X region, upstream of the large open reading frames (Fig. 4).

## DISCUSSION

We have sequenced a 3′ region of the HTLV-II genome of ≈1.6 kbp with an unknown function. HTLV-II resembles HTLV-I in a number of its properties, including biological functions, such as lymphoid target-cell specificity and T-cell transformation (1, 2, 6–10), and conservation of important structural features within the LTR (13). The X region represents another common structural feature that is present in the genomes of both HTLV types. The X region of HTLV-II has 61% sequence homology with the X region of HTLV-I, and shows homology as great as 75% in the region encompassing the 3′ two-thirds of the X region. The sequence conservation in this part of the X region in both types of HTLV strongly suggests that the X region serves an important function in virus replication and/or transformation. Three large open reading frames with overlapping sequences are present in the HTLV-II X region. If they began with initiation codons, these open reading frames would be sufficient to encode proteins of 15,100, 23,700, and 24,500 daltons. Although the predicted amino acid sequence of HTLV-II in these regions shows ≈60% homology with those of the X-II and X-III regions and 75% homology with that of the X-IV region of HTLV-I, the positions of initiation and termination codons would result in proteins of different predicted sizes.

Comparison of the LTR sequences of HTLV-II and HTLV-I indicates that these two viruses are only distantly related. It is likely that the two viruses evolved from a single ancestral virus and have retained common sequences that are important for replication (13). Of the corresponding open reading frames in the two viral genomes, Xc and X-IV share the greatest sequence homology, the longest stretch of contiguous codons uninterrupted by termination codons, and the greatest frequency of third-position differences relative to first- and second-position differences. Therefore, it is likely that Xc in HTLV-II and X-IV in HTLV-I encode functional proteins.

The high predicted amino acid homology and relatively high frequency of third-position differences holds true for 112 codons in Xc/X-IV located upstream of the first methionine codon. Therefore, it is probable that Xc is translated as a fused protein from a spliced mRNA. In this regard, it is interesting to note that a potential splice acceptor site is lo-

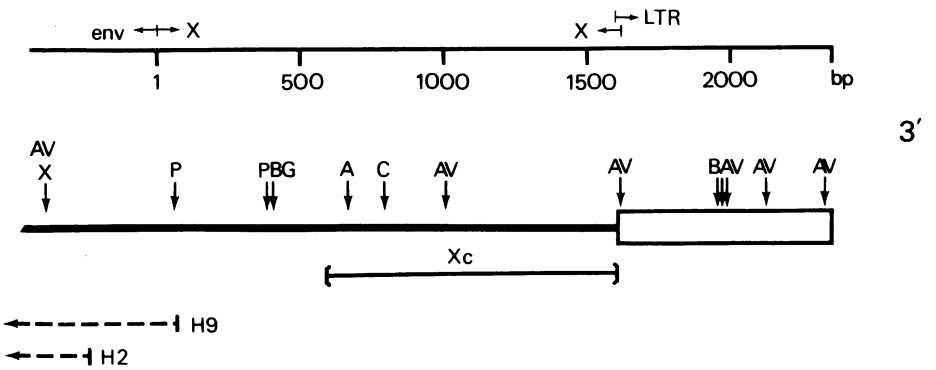


FIG. 4. Location of deletion endpoints in the HTLV-II H-2 and H-9 clones. The restriction enzyme map of the HTLV-II X region and 3′ LTR (represented by a box) is illustrated. Restriction enzyme sites are shown for *Acc* I (A), *Ava* I (AV), *Bam*HI (B), *Bgl* II (BG), *Cla* I (C), *Pst* I (P), and *Xho* I (X). Boundaries between *env* and X and between X and the LTR are indicated. The 3′ deletion endpoints of the defective H-2 and H-9 clones, as determined by restriction enzyme mapping and subsequent hybridization analysis (unpublished data) are shown in reference to the λH-6 restriction enzyme map. The location of Xc is denoted by a bracketed line. bp, Base pairs.

cated near the 5′ end of the Xc/X-IV region at nucleotide position 570.

Since animal retroviruses for which nucleic acid sequence information is available do not have sequences comparable to the X region of HTLV, it is likely that this X region has a unique function in viral replication and cellular transformation. The retention of the open reading frames in the X region of the deletion mutants of HTLV-II may be relevant to that region's potential function in transformation, particularly since the deletion mutants are present only in a subline of the Mo cells having much less stringent growth requirements than the parental Mo cells. Identification of the proteins encoded by the X region and X-region-specific mRNAs will be necessary to determine the significance of the X region.

**Note Added in Proof.** While this work was in press, Haseltine *et al.* (17) published a sequence of the 3′ region of HTLV-II that differs from that presented here at six nucleotide positions in the Xc region; four of these differences result in amino acid changes. These differences from our data may be due to sequence differences in the two provirus clones used for analysis; the significance, if any, of these differences must await demonstration of the infectivity of the cloned HTLV-II provirus used for sequencing (18).

We thank J. McLaughlin and S. Quan for technical assistance, J. Gassoh and T. Sekiya for comments on the manuscript, and T. Gojobori for helpful discussions. This work was partly supported by Grants-in-Aid for Cancer Research from the Ministry of Education, Science, and Culture and from the Ministry of Health and Welfare of Japan; U.S. Public Health Service Grants CA 30388, CA 32737, CA 09297, and CA 16042 awarded by the National Cancer Institute; and Grant PF-2182 from the American Cancer Society.

1. Saxon, A., Stevens, R. H. & Golde, D. W. (1978) *Ann. Intern. Med.* **88,** 323–326.
2. Kalyanaraman, V. S., Sarngadharan, M. G., Robert-Guroff, M., Miyoshi, I., Blayney, D., Golde, D. W. & Gallo, R. C. (1982) *Science* **218,** 571–573.
3. Poiesz, B. J., Ruscetti, F. W., Gazdar, A. F., Bunn, P. A., Minna, J. D. & Gallo, R. C. (1980) *Proc. Natl. Acad. Sci. USA* **77,** 7415–7419.
4. Hinuma, Y., Nagata, K., Hanaoka, M., Nakai, M., Matsumoto, T., Kinoshita, K.-I., Shirakawa, S. & Miyoshi, I. (1981) *Proc. Natl. Acad. Sci. USA* **78,** 6476–6480.
5. Blattner, W. A., Kalyanaraman, V. S., Robert-Guroff, M., Lister, T. A., Galton, D. A. G., Sarin, P. S., Crawford, M. H., Catovsky, D., Greaves, M. & Gallo, R. C. (1982) *Int. J. Cancer* **30,** 257–264.
6. Saxon, A., Stevens, R. H., Quan, S. G. & Golde, D. W. (1978) *J. Immunol.* **120,** 777–782.
7. Miyoshi, I., Kubonishi, I., Yoshimoto, S., Akagi, T., Ohtsuki, Y., Shiraishi, Y., Nagata, K. & Hinuma, Y. (1981) *Nature (London)* **294,** 770–771.
8. Yamamoto, N., Okada, M., Koyanagi, M. & Hinuma, Y. (1982) *Science* **217,** 737–739.
9. Popovic, M., Sarin, P. S., Robert-Guroff, M., Kalyanaraman, V. S., Mann, D., Minowada, J. & Gallo, R. C. (1983) *Science* **219,** 856–859.
10. Chen, I. S. Y., Quan, S. G. & Golde, D. W. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 7006–7009.
11. Seiki, M., Hattori, S., Hirayama, Y. & Yoshida, M. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 3618–3622.
12. Chen, I. S. Y., McLaughlin, J., Gasson, J. C., Clark, S. C. & Golde, D. W. (1983) *Nature (London)* **305,** 502–505.
13. Shimotohno, K., Golde, D. W., Miwa, M., Sugimura, T. & Chen, I. S. Y. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 1079–1083.
14. Maxam, A. M. & Gilbert, W. (1980) *Methods. Enzymol.* **65,** 499–560.
15. Chen, I. S. Y., McLaughlin, J. & Golde, D. W. (1984) *Nature (London)* **309,** 276–278.
16. IUPAC-IUB Commission on Biochemical Nomenclature (1968) *Eur. J. Biochem.* **5,** 151–153.
17. Haseltine, W. A., Sodroski, J., Patarca, R., Briggs, D., Perkins, D. & Wong-Staal, F. (1984) *Science* **225,** 419–421.
18. Gelmann, E. P., Franchini, G., Manzari, V., Wong-Staal, F. & Gallo, R. C. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 993–997.