host factors encoded by dominant alleles at the *Fv*-1 locus (*13*).

To our knowledge, these results are the first report of a viral capsid protein playing a critical role in the congenital transmission of a retrovirus. Whether capsid proteins affect the replication of other families of retroviruses in reproductive tissue is not known. However, since the ability to undergo efficient congenital transmission has survival value for exogenous but not endogenous viruses, the major capsid proteins for all exogenous and endogenous viruses may have undergone selection for their ability to ensure or restrict the replication of virus in reproductive tissue. If so, the capsid proteins of exogenous and endogenous viruses may provide genes that can be used to construct viruses that either will or will not undergo congenital transmission.

HARRIET L. ROBINSON
Worcester Foundation for Experimental Biology, Shrewsbury, Massachusetts 01545
ROBERT N. EISENMAN
Fred Hutchinson Cancer Research Center, Seattle, Washington 98104

References and Notes

1. H. Rubin, A. Cornelius, L. Fanshier, *Proc. Natl. Acad.. Sci. U.S.A.* **47**, 1058 (1962); B. R. Burmester, R. F. Gentry, N. F. Waters, *Poult. Sci.* **34**, 609 (1955).
2. R. M. Dougherty and H. S. DiStefano, *Cancer Res.* **27**, 322 (1967).
3. H. Rubin, L. Fanshier, A. Cornelius, W. F. Hughes, *Virology* **17**, 143 (1962).
4. H. L. Robinson and W. F. Lamoureux, *ibid.* **69**, 50 (1976); H. L. Robinson, S. M. Astrin, A. M. Senior, F. H. Salazar, *J. Virol.* **40**, 745 (1981).
5. H. L. Robinson, B. M. Blais, P. N. Tsichlis, J. M. Coffin, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1225 (1982).
6. H. L. Robinson, M. N. Pearson, D. W. DeSimone, P. N. Tsichlis, J. M. Coffin, *Cold Spring Harbor Symp. Quant. Biol.* **44**, 1133 (1979).
7. P. N. Tsichlis and J. M. Coffin, *J. Virol.* **33**, 238 (1980).
8. J. M. Coffin, P. N. Tsichlis, K. F. Conklin, A. Senior, H. L. Robinson, *Virology* **126**, 51 (1983).
9. R. J. Huebner, D. Armstrong, M. Okuyan, P. S. Sarma, H. C. Turner, *Proc. Natl. Acad. Sci. U.S.A.* **51**, 742 (1964).
10. R. A. Weiss, *J. Gen. Virol.* **5**, 511 (1969); *ibid.*, p. 529.
11. C. W. Rettenmier and H. Hanafusa, *J. Virol.* **24**, 850 (1977); R. Shaikh, M. Linial, J. Coffin, R. Eisenman, *Virology* **87**, 326 (1978).
12. J. Schindler, R. Hynes, N. Hopkins, *J. Virol.* **23**, 700 (1977); D. V. Faller and N. Hopkins, *ibid.* **26**, 153 (1978); J. W. Gautsch, J. H. Elder, J. Schindler, F. C. Jensen, R. A. Lerner, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 4170 (1978).
13. T. Pincus, J. W. Hartley, W. P. Rowe, *J. Exp. Med.* **133**, 1219 (1971); F. Lilly and T. Pincus, *Adv. Cancer Res.* **17**, 231 (1973).
14. R. N. Eisenman and V. M. Vogt, *Biochim. Biophys. Acta* **473**, 187 (1978).
15. H. L. Robinson, unpublished observation.
16. We thank B. Blais, G. Gagnon, and S. Johnson for technical assistance; H. Hanafusa and T. Hanafusa for the RAV-60 viruses; J. Coffin and P. Tsichlis for NTRE-7 and NTRE-2; P. Tsichlis for pRAV-0; G. Payne, M. Bishop, and H. Varmus for pRAV-1; and D. Steffen and S. Wadsworth for comments on the manuscript. Supported by NIH grants R01 CA23086 and P30 12708 and by the W. J. Tannenberg fund (H.L.R.) and NIH grant R01 CA20525 (R.N.E.). R.N.E. is a scholar of the Leukemia Society of America.

12 March 1984; accepted 4 May 1984

Structure of 3′ Terminal Region of Type II Human T Lymphotropic Virus: Evidence for New Coding Region

Abstract. *The sequence of the 3′ terminus of the human T lymphotropic virus type II (HTLV-II) was determined and compared to the corresponding sequence of HTLV-I. The 1557-nucleotide-long sequence can be divided into a 5′ region that is not conserved between the two viruses, and a 3′, 1011-nucleotide-long region that is highly conserved and that corresponds precisely with a long open reading frame for both HTLV-I and -II. The proteins that could be encoded by these open reading frames have a molecular weight of about 38,000 and are closely related in primary amino acid sequence. The genomic structure in the 3′ region of HTLV was found to be similar to that of bovine leukemia virus.*

The human T lymphotropic viruses (HTLV) are a family of retroviruses that are associated with T-cell abnormalities (*1*). Isolates known as HTLV-I are associated with an aggressive form of adult T-cell leukemia or lymphoma (*1*). An infrequent isolate known as HTLV-II was first identified in a patient with a T-cell variant of hairy cell leukemia (*2*). Recently, some viruses collectively called HTLV-III were isolated from patients with the acquired immune deficiency syndrome (*3*).

The genomes of HTLV-I and -II differ from those of the nonacute retroviruses, which encode only the *gag*, *pol*, and *env* genes, in that they have an additional sequence that is approximately 1600 nucleotides long. This sequence is located between the 3′ end of the *env* gene and the 5′ end of the U3 region of the proviral long terminal repeat (LTR) (*4*).

Although this sequence occupies a position similar to the *src* gene in Rous sarcoma virus, it is not homologous to conserved mammalian genes and therefore differs from the oncogenes of transforming retroviruses (*4*). There is some evidence that this region contains a functional gene. Heteroduplex analysis of HTLV-I and -II reveals a conserved sequence about 1000 nucleotides long near the 3′ terminus of the genome (*5*). Spliced messenger RNA (mRNA) species that contain sequences that are unique to the 5′ end of the viral genome (U5 LTR sequences) and a portion of the 3′ sequence are observed in HTLV-infected cells and in some fresh tumor cells (*6*). Seiki *et al.* (*4*) note that several open reading frames occur within the 3′ sequence of HTLV-I.

To obtain a clearer understanding of the potential role of the 3′ region of HTLV, we determined the primary nucleotide sequence of the region located between the 3′ end of the *env* gene and the LTR of a cloned HTLV-II provirus, MO15A (*7*).

The nucleotide sequence of 1557 bases of the 3′ terminal region of HTLV-II is presented in Fig. 1. This sequence can be divided into two regions. One region, 546 nucleotides long, is located at the 5′ end of the sequence and has either no or very little similarity to the corresponding sequences in HTLV-I. For this reason we call this sequence the nonconserved region (NCR). A second region, 1011 nucleotides long, comprises the 3′ portion of this sequence. This sequence is very similar to that of HTLV-I and is identical at 765 of 1011 nucleotides (76 percent identity).

A new gene? The perimeters of the 1011 nucleotide sequence of the HTLV-II genome correspond precisely with a single long open reading frame capable of encoding a polypeptide 337 amino acids long. A corresponding sequence of HTLV-I also encompasses a single long open reading frame capable of encoding a polypeptide 357 amino acids long. We call the nucleotide sequence containing these long open reading frames the LOR region (nucleotides 566 to 1557 in HTLV-II) (Fig. 1).

The predicted amino acid sequences of both polypeptides are presented in Fig. 1. The potential proteins encoded by the LOR regions of HTLV-I and -II are of approximately the same length and are identical in 259 of 337 of the amino acids (77 percent identity). The degree of similarity of these two proteins is even more striking if conservative amino acid substitutions are considered (89 percent similar). The distribution of hydrophilic and hydrophobic regions of these proteins is remarkably similar (Fig. 2).

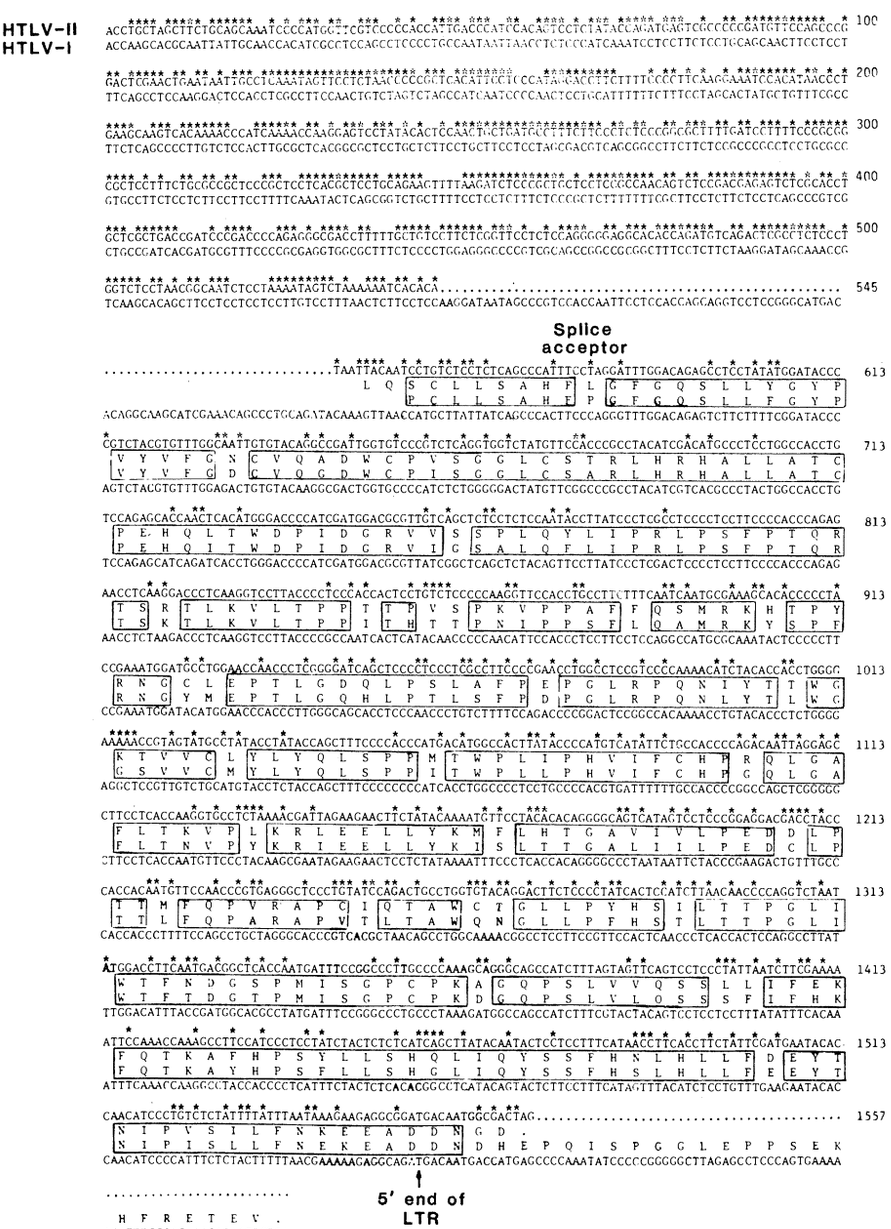
We also note the existence of a splice acceptor consensus sequence located at the 5′ end of the open reading frame (Fig. 1). Although no methionine codon occurs at the 5′ end of the open reading frames of HTLV-I and -II, a fusion protein synthesized from a spliced mRNA can be envisioned. Several other splice acceptor sequences occur within this reading frame from which smaller fusion proteins might also be made.

These observations suggest that the 3′ terminal region of HTLV contains a new gene that encodes a protein with a mo-

lecular weight of at least 38,000. Such a protein could be translated from the 2.2-kb spliced mRNA species containing LOR sequences found in HTLV-infected cells (*6*). A protein of molecular weight 38,000 to 42,000 in HTLV-I–infected cell lines has been noted that is recognized by the serum of persons infected with HTLV-I, but not by serum from control subjects (*8*).

Several other open reading frames exist in the region between the *env* gene and the LTR of both HTLV-I and -II. Seiki *et al.* (*4*) have identified four such regions, pX I to pX IV. The pX IV region corresponds to the carboxyl terminus of the peptide that could be encoded by the LOR region. No region of predicted protein similarity could be found in HTLV-II that corresponds to pX I or pX III. A further argument against the functional importance of pX I is that an 11-nucleotide deletion that destroys the pX I open reading frame occurs in an HTLV-Ic isolate with apparently complete biological activity (*9*). Another open reading frame in the LOR region of HTLV-II (nucleotides 530 to 1325) includes a region exhibiting 65 percent amino acid homology to pX II. The significance of this similarity is not clear, because the pX II peptide is much shorter than the corresponding peptide in HTLV-II (87 compared to 265 amino acids). Sequence similarity here could arise as a result of conservation of the LOR protein in the other open reading frame.

We have also reported (*8*) that *trans*-acting factors, either directly encoded by



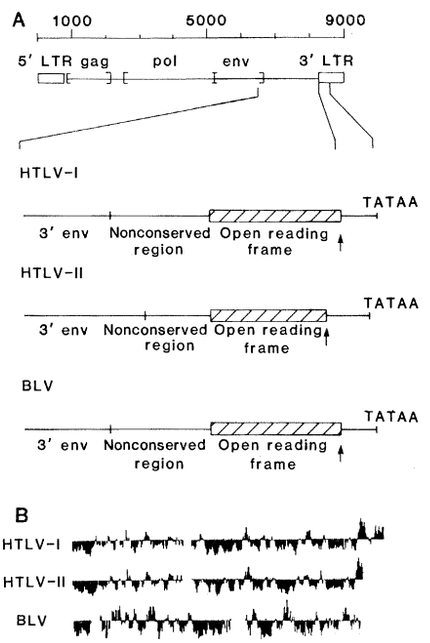


Fig. 1 (left). Nucleotide sequence of the HTLV-II 3′ terminal region and predicted amino acid sequence of its potential product. Plasmid DNA containing the 3′ portion of MO15A, an HTLV-II proviral clone (*7*), was cleaved with either Cla I or Bgl II, which cut the plasmid uniquely at a single site. After timed digestion with Bal 31 exonuclease, the ends were blunted with T4 DNA polymerase and synthetic linkers were added prior to recloning. Linker sites separated by increments of 100 to 200 nucleotides were end-labeled and the fragments sequenced by the method of Maxam and Gilbert (*19*). The sequence of the 3′ region following the termination codon for the envelope gene is presented for HTLV-II and HTLV-I. The HTLV-II sequence is numbered according to the nucleotides following the envelope stop codon. Asterisks represent differences between the DNA sequences. The positions of a conserved splice acceptor consensus sequence and the 5′ end of the LTR are noted. Note that the sequence is not well conserved 5′ to the putative splice acceptor site but is very well conserved 3′ to this site. The latter sequence corresponds to a long open reading frame (LOR) region. The predicted amino acid sequences of the potential products of the HTLV-I and HTLV-II 3′ open reading frames are optimally aligned. Boxed regions indicate amino acid identity or conservative amino acid substitutions between the sequences. Fig. 2 (right). The open reading frames of the HTLV and BLV genomes. (A) The position of 3′ open reading frames in the genomes of HTLV-I and -II and of BLV. The 3′ end of the envelope gene is shown, as well as the 5′ terminus of the LTR (↑) and the promoter (TATAA) sequence. The positions of the nonconserved regions and the open reading frames (hatched boxes) are displayed. (B) The relative hydrophilicity of the 3′ open reading frame products of HTLV-I, HTLV-II, and BLV calculated according to the method of Hopp and Woods (*20*). Hydrophilic regions are shown above the axis, hydrophobic regions below. Dotted lines represent gaps introduced to maintain maximal alignment of protein sequence.

HTLV or induced by HTLV infection, substantially augment gene expression directed by HTLV LTR sequences. The phenomenon of *trans*-activation distinguishes HTLV from other retroviruses. The unusual structure of the 3′ terminus of HTLV also distinguishes these from most other retroviruses. For this reason, we suggest that the protein encoded by the LOR region may mediate transcriptional changes observed in HTLV-infected cells. In this regard, we note that transcription directed by the HTLV-I LTR is activated to high levels in a cell line, C81-66, that expresses the 42,000-dalton HTLV-I–associated protein but not HTLV *gag*, *pol*, or *env* products (*8*). We further suggest that the HTLV LOR product mediates both the *trans*-activating and transforming effects of HTLV infection. We note that *trans*-acting transcriptional activities have been associated with the transforming genes of other tumor viruses, notably adenovirus and SV40 (*10, 11*). The existence of a potential transforming function within the HTLV genome may explain the ability of the virus to transform cells in vitro, as well as the absence of specific integration sites in tumor cells and the absence of chronic viremia in target tissues (*12–14*). Such a transforming function would differ from that of other retroviruses because, unlike the oncogenes, the sequence that encodes the putative transforming gene will not anneal to the highly conserved cellular sequences (*4*).

Comparison with the bovine leukemia virus genome. We noticed that the 3′ genome of another retrovirus, bovine leukemia virus (BLV), also contains an LOR frame located 3′ to the envelope glycoprotein gene that could encode a protein of a size similar to that of HTLV (*15, 16*) (Fig. 2). There is evidence for the existence of a subgenomic spliced mRNA species that contains the 3′ open reading frame but not the *gag*, *pol*, and *env* gene sequences in BLV-producing cell lines (*17*).

Although the similarity in structure of the HTLV and BLV proteins is insufficient to indicate that they have a common functional role, the overall similarity in genomic structure, including the location of a 5′ NCR and 3′ LOR frame, and the previously described similarity in protein antigenicity of the two viruses (*1, 14*) suggests that they are functionally similar. Moreover, there is a similarity in the distribution of hydrophobic and hydrophilic regions of the HTLV and BLV polypeptides. We note that the disease induced by BLV has characteristics similar to those associated with HTLV-I, namely, a long latent period sometimes preceded by persistent lymphocytosis, an absence of chronic viremia in target organs preceding disease, and an absence of preferred integration sites in tumor cells (*18*). These features could be expected of viruses that contain an LOR product mediating transformation.

The biology, structure, and pathology of HTLV and BLV differ from other transforming retroviruses such that we propose that they be considered a new subgroup of retroviruses distinct from both the nonacute transforming viruses that contain only the *gag*, *pol*, and *env* genes and the acute transforming viruses that encode oncogenes.

WILLIAM A. HASELTINE
Dana-Farber Cancer Institute and
Department of Pathology,
Harvard Medical School, and
Department of Cancer Biology,
Harvard School of Public Health,
Boston, Massachusetts 02115
JOSEPH SODROSKI
ROBERTO PATARCA
DEBRA BRIGGS
DENNIS PERKINS
Dana-Farber Cancer Institute
FLOSSIE WONG-STAAL
Laboratory of Tumor Cell Biology,
National Cancer Institute,
Bethesda, Maryland 20205

References and Notes

1. B. J. Poiesz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7415 (1980); V. S. Kalyanaraman, M. Sargadharan, P. Bunn, J. Minna, R. C. Gallo, *Nature (London)* **294**, 271 (1981); M. Robert-Guroff *et al.*, *J. Exp. Med.* **154**, 1957 (1981); M. Yoshida, I. Miyoshi, Y. Hinuma, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 2031 (1982).
2. V. S. Kalyanaraman *et al.*, *Science* **218**, 571 (1982).
3. M. Popovic, M. G. Sarngadharan, E. Read, R. C. Gallo, *Science* **224**, 497 (1984); R. C. Gallo *et al.*, *ibid.*, p. 500; J. Schüpbach *et al.*, *ibid.*, p. 503; M. Sarngadharan, M. Popovic, L. Bruch, J. Schüpbach, R. C. Gallo, *ibid.*, p. 506.
4. M. Seiki, S. Hattori, Y. Hirayama, M. Yoshida, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 3618 (1983).
5. G. M. Shaw *et al.*, *ibid.*, in press.
6. G. Franchini, F. Wong-Staal, R. C. Gallo, in preparation.
7. E. P. Gelman, G. Franchini, V. Manzari, F. Wong-Staal, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 993 (1984).
8. J. Sodroski *et al.*, *Science* **225**, 421 (1984); T. H. Lee *et al.*, in preparation.
9. L. Ratner and F. Wong-Staal, in preparation.
10. J. Brady, J. B. Bolen, M. Radonovich, N. Salzman, G. Khoury, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2040 (1984).
11. N. Jones and T. Shenk, *ibid.* **76**, 3665 (1979); J. R. Nevins, *Cell* **26**, 213 (1981); A. Berk, F. Lee, T. Harrison, J. Williams, P. A. Sharp, *ibid.* **17**, 935 (1979); R. B. Gaynor. D. Hillman, A. Berk, *Proc. Natl. Acad. Sci. U.S.A.* **81** 1193 (1984).
12. I. Miyoshi *et al.*, *Nature (London)* **294**, 770 (1981); M. Popovic, G. Lange-Wantzin, P. S. Sarin, D. Mann, R. C. Gallo, *Proc. Natl. Acad. Sci. U.S.A.* **80**, 5402 (1983); N. Yamamoto, M. Okada, Y. Koyanagi, M. Kannagi, Y. Hinuma, *Science* **217**, 737 (1982).
13. B. Hahn *et al*, *Nature (London)* **305**, 340 (1983).
14. R. C. Gallo and F. Wong-Staal, *Blood* **60**, 545 (1982).
15. A. Burny, personal communication.
16. N. R. Rice *et al.*, *Virology*, in press.
17. J. Ghysdael, R. Kettmann, A. Burny, *J. Virol.* **29**, 1087 (1979).
18. J. F. Ferrer, D. A. Abt, D. M. Bhatt, R. R. Marshak, *Cancer Res.* **34**, 893 (1974); P. S. Paul *et al.*, *Am. J. Vet. Res.* **38**, 873 (1977); R. Kettmann *et al.*, *J. Virol.* **47**, 146 (1983); D. Gregoire *et al.*, *ibid.* **50**, 275 (1984); J. Deschamp, R. Kettman, A. Burny, *J. Virol.* **40**, 605 (1981); R. C. Gallo *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 5680 (1982).
19. A. Maxam and W. Gilbert, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 564 (1977).
20. T. P. Hopp and K. R. Woods, *ibid.* **78**, 3824 (1981).
21. We thank A. Burny for providing the sequence of the 3′ region of the BLV, N. R. Rice *et al.* for a preprint of their manuscript, and R. C. Gallo, M. Essex, and J. Coffin for helpful discussions. Supported by an American Cancer Society Directors Grant. J.G.S. was supported by NIH postdoctoral fellowship CA07094.

31 May 1984; accepted 21 June 1984

Sequence of the Envelope Glycoprotein Gene of Type II Human T Lymphotropic Virus

Abstract. *The sequence of the envelope glycoprotein gene of type II human T lymphotropic virus (HTLV) is presented. The predicted amino acid sequence is similar to that of the corresponding protein of HTLV type I, in that the proteins share the same amino acids at 336 of 488 residues, and 68 of the 152 differences are of a conservative nature. The overall structural similarity of these proteins provides an explanation for the antigenic cross-reactivity observed among diverse members of the HTLV retrovirus family by procedures that assay for the viral envelope glycoprotein, for example, membrane immunofluorescence.*

Human T-cell leukemia viruses have been implicated as the etiological agents of several human diseases. The most prevalent type, HTLV-I, is associated with a high incidence of an aggressive form of adult T-cell leukemia (ATLL) and several unusual forms of mycosis fungoides and Sezary syndrome (*1*). A second member of the family, HTLV-II, has been isolated from a patient with benign hairy cell leukemia of T-cell origin (*2*). Recently, a new group of viruses, HTLV-III, was isolated from patients with acquired immune deficiency syndrome (AIDS) (*3*).

The envelope glycoprotein is the major antigen recognized by the serum of persons infected with HTLV (*4*). In this respect HTLV resembles several other retroviruses for which the envelope glycoprotein is typically the most antigenic viral polypeptide (*5*). Moreover, most