Expression of the *pX* Gene of HTLV-I: General Splicing Mechanism in the HTLV Family

Abstract. *Human T-cell leukemia virus type I (HTLV-I) is an etiological agent of adult T-cell leukemia. A viral gene* pX *encodes for p40^x and it has been proposed that this protein* trans*-activates the viral long terminal repeat and possibly some cellular genes; this activation may be associated with T-cell transformation. The mechanism of* pX *gene expression and the primary structure of p40^x are now reported. Two-step splicing generates the 2.1-kilobase* pX *mRNA; the initiator methionine for* env *becomes part of the* pX *protein. These splicing signals are conserved among all members of the HTLV family except for the acquired immune deficiency syndrome–associated viruses.*

Human T-cell leukemia virus type I (HTLV-I) was the first isolated member of the HTLV family (*1–3*). It is closely associated with adult T-cell leukemia (ATL) (*2–4*), which is endemic in southwest Japan (*5*), the West Indies (*6*), and Africa (*7*). Infection of a target cell with this virus is a prerequisite for development of ATL (*8*).

Nucleotide sequence analysis of the HTLV-I genome indicated unusual structural features on the basis of which HTLV-I was classified into a new group of retroviruses (*9, 10*). These features include an unusually long R sequence (a repeated sequence that is found at both ends of viral RNA genome) and a potential novel secondary structure in the R region. The large stem and loop structure in the R region may function in the efficient termination of transcription and polyadenylation (*10*). Another unusual structure is an extra sequence (*pX*) between *env* and the long terminal repeat (LTR) that has the capacity to code for a protein of 40 kilodaltons (kD) (*10*). Other members of the HTLV family, namely HTLV-II (*11*), bovine leukemia virus (BLV) (*12*), and simian T-cell leukemia virus (STLV) (*13*), also have these unusual structural features of the LTR and *pX* regions. This conservation of the R and *pX* regions among distantly related viruses of the HTLV family suggests that these regions have important functions in viral replication or pathogenicity.

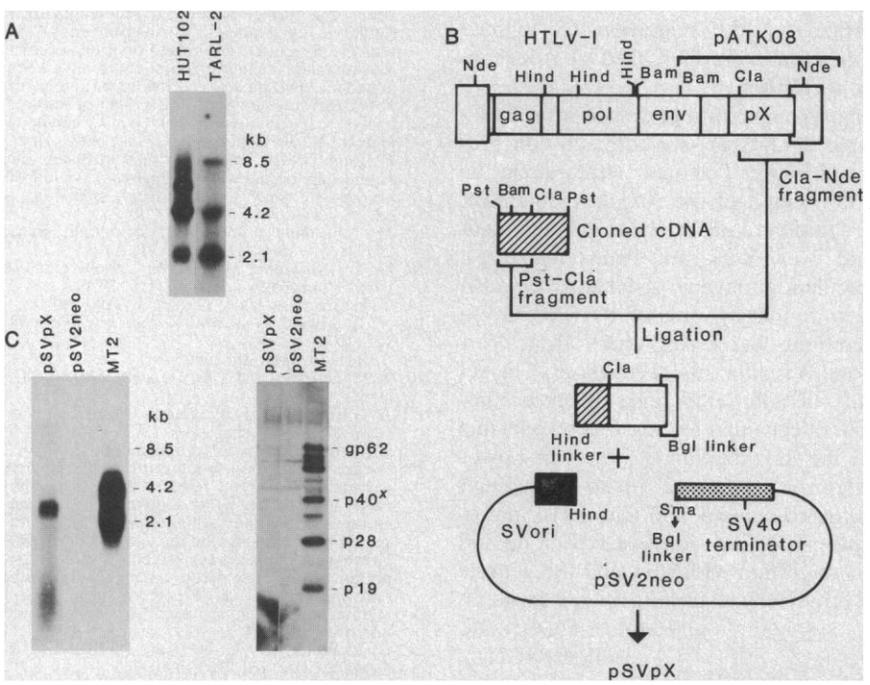

Fig. 1. Construction of plasmid for p40^x expression. (A) Detection of subgenomic *pX* mRNA in HTLV-I–infected cells. Cytoplasmic mRNA was isolated from human T-cell line HUT102 and rat T-cell line TARL-2 and fractionated by formaldehyde-agarose gel electrophoresis. The viral RNA was detected with a ^{32}P-labeled HTLV-I probe. (B) Construction of pSVpX for *pX* gene expression. The complete cDNA clone was reconstructed from a cloned cDNA with defects in the 3′-terminal region, and a provirus clone pATK08 (*10*). The reconstructed cDNA was inserted into the SV40 expression vector. (C) Expression of p40^x. COS-7 cells were transfected (*28*) with 10 μg of pSVpX, or with 10 μg of pSV2neo as a control. After incubation for 48 hours, cells were collected and RNA (left) and proteins (right) were analyzed. (Left) Cytoplasmic RNA was analyzed by the blotting procedure as in (A) but with *pX*-specific probe. (Right) Transfected cells were lysed and sonicated in RIPA buffer (50 m*M* Tris-HCl, *p*H 7.5, 150 m*M* NaCl, 0.1 percent sodium dodecyl-sulfate, 1 percent Triton X-100, and 1 percent sodium deoxycholate) and analyzed by blot-hybridization with sera from ATL patients and ^{125}I-labeled antibody to human immunoglobulin. As control, an HTLV-producing human T-cell line, MT2, was included. Nde, NdeI; Hind, Hind III; Bam, Bam HI; Cla, Cla I; Pst, Pst I; Bgl, Bgl I.

Recently, a 40-kD product of the *pX* gene, p40^x, was identified by means of specific antibodies to the COOH-terminal region of the protein predicted from the *pX* nucleotide sequence (*14*). Moreover, p40^x was suggested to enhance transcription initiated at its own LTR in a *trans*-acting manner (*15*). This finding suggested that p40^x is also involved in leukemogenesis by activating certain cellular genes whose regulation is similar to the LTR. We have previously excluded the possibility of *cis*-function of the LTR of the integrated provirus genome because the site of integration of the provirus in tumor cells varied among different ATL patients (*16*). This *trans*-acting viral protein could be the *pX* gene product, p40^x. The structure of p40^x could not be predicted from the genomic sequence, because the open reading frame in the *pX* sequence coding for p40^x lacks an initiation codon (ATG) at its 5′ end, although the frame has the capacity to encode the 40-kD protein. Therefore, p40^x was suggested to be encoded by spliced messenger RNA (mRNA) (*14*). Information on the mechanism of mRNA formation and the NH_2-terminal structure of p40^x would be useful for understanding regulation of *pX* expression in infected cells and the function of p40^x. We now describe cloning of complementary DNA (cDNA) and structural analysis of p40^x mRNA and propose a general mechanism for viral gene expression in the HTLV family.

A rat T-cell line, TARL-2 (*17*), was used as a source of mRNA, because this cell line has one copy of the intact provirus genome in contrast to other human T-cell lines, which contain multiple copies of the proviruses including defective ones (*3, 18*). Cytoplasmic, polyadenylated RNA was isolated by oligo(dT)-cellulose column chromatography and analyzed by blot-hybridization. Three species of viral RNA were detected with a representative HTLV-I probe (Fig. 1A). They were 8.5, 4.2, and 2.1 kilobases (kb) in size and were concluded to be genomic RNA and subgenomic mRNA's for *env* and *pX*, respectively, by their sizes and gene-specific hybridization. The RNA's were separated by centrifugation in a sucrose gradient, and fractions containing 2.1-kb mRNA were collected and used as template for com-

plementary DNA (cDNA) synthesis. Oligo(dC)-tailed double stranded cDNA was synthesized by the method of Land *et al.* (*19*) and annealed to oligo(dG)-tailed pBR322 at the Pst I site. The DNA was transfected into the MC 1061 strain of *Escherichia coli* (*20*), and colonies were screened by in situ hybridization with the *pX* region of the HTLV-I as a probe. Five positive colonies were obtained from 4×10^4 transformants and one plasmid, containing the largest insert (1.5 kb), was analyzed further. Initial restriction enzyme analysis showed that this clone contained the 5′ region of mRNA but had lost about 500 bases of the 3′ portion of the coding frame. This defect in the 3′-portion should be due to the incomplete DNA synthesis of the second strand.

To confirm that this cDNA clone is functionally active for the expression of $p40^x$, the defective 3′ portion of the cDNA was first replaced with the sequence containing a complete 3′ portion isolated from the proviral clone pATK08 (*10*) (Fig. 1B), and then the constructed cDNA was joined in place of the *neo* gene in the pSV2neo (*21*) (Fig. 1B). The resultant plasmid, pSVpX, was transfected into COS-7 cells and the transient expression of the *pX* gene was analyzed (Fig. 1C). A single band of the mRNA containing the *pX* sequence was detected in cells transfected with pSVpX, but not in those infected with pSV2neo. The size of the mRNA (3.1 kb) in the transfectant was different from that in the HTLV-I–infected T-cell line. This was expected as the transcription of the *pX* gene in the plasmid is terminated by the SV40 termination signal instead of the LTR sequence, and so the mRNA should be slightly larger than the original mRNA. The 40-kD protein was also detected in pSVpX transfectants with serum from a patient who had ATL (Fig. 1C) or with rabbit antibodies against synthetic peptide corresponding to the COOH-terminal portion of $p40^x$. These results indicate that the cDNA clone contained enough information to express $p40^x$.

Comparison of the cDNA sequence with that of the provirus genome revealed that the cDNA is composed of three blocks (Fig. 2A). The first block consists of 118 nucleotides derived from the R region (from the cap site at position 354 to position 471) of the LTR. This spliced position at 471 is followed by the consensus sequence (*22*) for the splicing donor site AGGTAAG at positions 470 to 476 in the R region. The second block consists of 191 nucleotides (position 4993 to 5183) in the *pol* region. The junction sites at the two ends of this block correspond to the splicing acceptor sequence TATTTCAAG (4984 to 4992) and donor sequence GGGTAAG (5182 to 5188) in the proviral DNA. The initiation codon ATG (5180 to 5182) for the *env* gene (*10, 23*) is located just 5′ to the donor sequence. The third block consists of all of the *pX* sequence starting from position 7032, which is preceded by the splicing acceptor site TATTATCAG (position 7293 to 7301). Presence of the splicing site at this position was previously demonstrated by S_1 nuclease analysis (*24*). As a result of this second splicing, the *pX* reading frame (7032 to 8356), which can

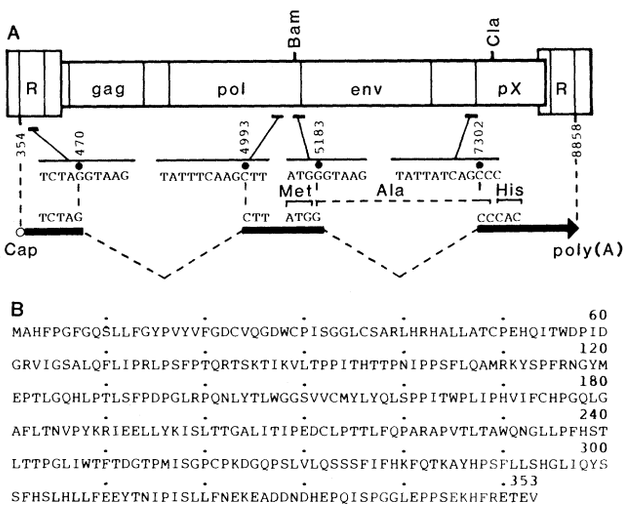

Fig. 2 (left). Summary of nucleotide sequence analysis of the cDNA clone of *pX* mRNA. (A) Splicing sites for mRNA formation. The cDNA, composed of three contiguous segments as shown by thick lines, and its correspondence to the proviral genome are shown. The third domain, the open reading frame from the *pX* region, was sequenced from position 7302 to 7950, which was the 3′ end of the cDNA clone. The sequence determined was identical to that of the provirus genome reported previously except for the following base replacements: T to A at position 5171, G to A at 7374, G to C at 7725, and G to A at 7801. (B) The complete amino acid sequence of $p40^x$ was deduced from the nucleotide sequence of the cDNA clone. The calculated molecular weight of the protein is 39,482 daltons. The following abbreviations were used for amino acids: A, alanine; C, cysteine; D, aspartic acid; E, glutamic acid; F, phenylalanine; G, glycine; H, histidine; I, isoleucine; K, lysine; L, leucine; M, methionine; N, asparagine; P, proline; Q, glutamine; R, arginine; S, serine; T, threonine; V, valine; W, tryptophan; Y, tyrosine.

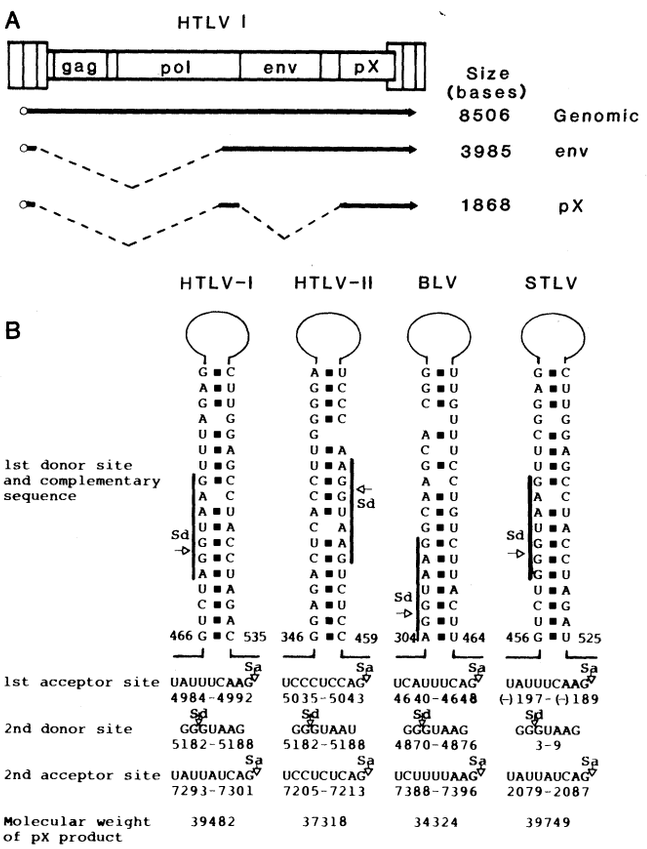

Fig. 3 (right). Conservation of splicing signals in the HTLV family. (A) Splicings to generate the subgenomic mRNA of HTLV. Thick lines represent RNA segments transcribed from the provirus genome and dotted lines indicate portions spliced out from the genomic RNA. (B) Conserved sequences for characteristic splicings in HTLV family members. Possible donor and acceptor sites in HTLV family members, HTLV-I, HTLV-II, BLV, and STLV are compared. For identification of the splicing sites, bases are numbered from the first base of the 5′-LTR except for the second splicing of STLV, where the bases are numbered from the first base of the *env* gene (as the total nucleotide sequence of the genome is not known). Bars beside stems of the stem and loop structure indicate the consensus sequences (*22*) for the splicing donor sites; arrows with Sd and Sa indicate splicing donor and acceptor sites, respectively.

code for 352 amino acids, is joined to the ATGG (5180 to 5183) of the *env* gene so that ATG is aligned in the frame. Thus, the initiation codon ATG for *env* is used for initiation of $p40^x$ translation. The other four cDNA clones showed similar restriction maps although they are also defective, supporting the conclusion that the sequenced cDNA clone represents the majority of the *pX* mRNA population. This conclusion was also consistent with the observation that the 2.1-kb mRNA did not significantly hybridize with the U5 probe. Similar splicing was also found in HTLV-II by Wachsman *et al.* (see *24a*).

The structure of the cDNA clone showed that subgenomic mRNA for $p40^x$ is formed by two-step splicing (Fig. 3A) and that one of the splice donor sites is located in the R region of LTR. This feature is unique to HTLV and might relate to the unusually long R region. We have proposed (*10*) that the ability of the long R sequence to form a secondary structure at the 3′ end of viral mRNA plays a role in transcriptional termination. In addition to this, the R sequence was suggested to regulate mRNA splicing expressing the *env* and *pX* genes, because a complementary sequence to this splicing donor site is found in the R region just after this donor site. Thus the donor site can form a stem structure with 18 bases (Fig. 3B). The secondary structure of the transcript at this splicing donor site would allow this region to compete with U1 RNA in a small nuclear ribonucleoprotein complex; the U1 complex is involved in exact RNA splicing (*25*). Alteration of the pairing ratio of U1 RNA to the stem and loop structure may control the splicing in the R region, eventually affecting the expression of the *env* and *pX* genes. For specific regulation of the *pX* expression, other mechanisms may be required.

One of the splicing acceptor sites is located in the *pol* region, 187 bases upstream from the ATG for the *env* gene. This structure indicates that the subgenomic RNA generated by single splicing is *env* mRNA (Fig. 3A). The calculated size of this spliced product, 3985 bases without the polyadenylated stretch, is consistent with that of one of the subgenomic mRNA's detected in infected cell lines. Probably, the other splicing takes place on the *env* mRNA and produces *pX* mRNA coding for $p40^x$. In this mRNA, the initiation codon, ATG for the *env*, is used to initiate the translation of $p40^x$, and only the first methionine is brought onto $p40^x$ from the *env* domain. Thus the molecular weight of the *pX* gene product was calculated as 39,482.

The key sequences for splicing found in HTLV-I are also present at corresponding positions in HTLV-II (*11, 26*), BLV (*12*), and STLV (*13*) as shown in Fig. 3, suggesting that the unusual splicing mechanisms producing the *env* and *pX* mRNA's are common to all four viruses. A minor exception occurs in BLV; the initiation codon for *pX* translation is located 44 bases downstream from that of the *env* gene, out of the frame. Thus, the *pX* and *env* genes do not share the same initiation codon. However, AIDS-associated viruses (*27*) are different from the other members of HTLV family in the organization of these critical sequences. Thus, a mechanism of the gene expression of AIDS-associated viruses seem to be different from the others.

MOTOHARU SEIKI
ATSUKO HIKIKOSHI
Department of Viral Oncology, Cancer Institute, Kami-Ikebukuro, Toshima-ku, Tokyo 170, Japan
TADATSUGU TANIGUCHI
Institute for Molecular and Cellular Biology, Osaka University, Suita-shi, Osaka 565, Japan
MITSUAKI YOSHIDA
Department of Viral Oncology, Cancer Institute

References and Notes

1. B. J. Poiesz *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 7415 (1980); M. R. Reitz *et al.*, *ibid.* **78**, 1887 (1981).
2. Y. Hinuma *et al.*, *ibid.*, p. 6476.
3. M. Yoshida, I. Miyoshi, Y. Hinuma, *ibid.* **79**, 2031 (1982).
4. M. Robert-Guroff *et al.*, *Science* **215**, 975 (1982); V. S. Kalyanaraman *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1653 (1982).
5. Y. Hinuma *et al.*, *Int. J. Cancer* **29**, 631 (1982).
6. W. A. Blattner *et al.*, *ibid.* **30**, 257 (1982); J. D. Schüpback, *et al.*, *Cancer Res.* **43**, 886 (1983).
7. W. Saxinger *et al.*, *Science* **225**, 1473 (1984).
8. M. Yoshida *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 2534 (1984).
9. M. Seiki, S. Hattori, M. Yoshida, *ibid.* **79**, 6899 (1982).
10. M. Seiki *et al.*, *ibid.* **80**, 3618 (1983).
11. K. Shimotohno *et al.*, *ibid.* **81**, 6657 (1984); W. A. Haseltine *et al.*, *Science* **225**, 419 (1984).
12. N. R. Rice *et al.*, *Virology* **138**, 82 (1984); N. Sagata *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
13. A. Komuro *et al.*, *Virology* **138**, 373 (1984); T. Watanabe *et al.*, *ibid.*, in press; H.-G. Guo, F. Wong-Staal, R. C. Gallo, *Science* **223**, 1195 (1984).
14. T. Kiyokawa *et al.*, *GANN* **75**, 747 (1984); M. Miwa *et al.*, *ibid.*, p. 751; T. H. Lee *et al.*, *Science* **226**, 57 (1984); D. J. Slamon *et al.*, *ibid.*, p. 61.
15. J. G. Sodroski, C. A. Rosen, W. A. Haseltine, *Science* **225**, 381 (1984); J. Fujisawa *et al.*, *Proc. Natl. Acad. Sci. U.S.A.*, in press.
16. M. Seiki, R. Eddy, T. B. Shows, M. Yoshida, *Nature (London)* **309**, 640 (1984).
17. M. Tateno *et al.*, *J. Exp. Med.* **159**, 1105 (1984).
18. T. Watanabe, M. Seiki, M. Yoshida, *Virology* **133**, 238 (1984).
19. H. Land *et al.*, *Nucleic Acids Res.* **9**, 2251 (1981).
20. M. Casadaban and S. M. Cohen, *J. Mol. Biol.* **138**, 179 (1980).
21. P. J. Southern and P. Berg, *J. Mol. Appl. Genet.* **1**, 327 (1982).
22. P. A. Sharp, *Cell* **23**, 643 (1984).
23. T. H. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **81**, 3856 (1984).
24. W. Wachsman *et al.*, *Science* **226**, 177 (1984).
24a. W. Wachsman *et al.*, *ibid.* **228**, 1534 (1985).
25. M. R. Lerner *et al.*, *Nature (London)* **283**, 220 (1980); J. Rogers and R. Wall, *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1877 (1980).
26. K. Shimotohno *et al.*, *ibid.*, in press.
27. L. Ratner *et al.*, *Nature (London)* **313**, 277 (1985); R. Sanchez-Pescador *et al.*, *Science* **227**, 484 (1985); M. A. Muesing *et al.*, *Nature (London)* **313**, 450 (1985); S. Wain-Hobson *et al.*, *Cell* **40**, 9 (1985).
28. P. Mellon *et al.*, *Cell* **27**, 279 (1981).
29. We thank G. Yamada, Institute for Molecular and Cellular Biology, Osaka University, for his useful discussion and help. Supported in part by a Grant-in-Aid for Special Project Research, Cancer-Bioscience; a Grant-in-Aid for Cancer Research, from the Ministry of Education, Science and Culture of Japan; and by a Research Grant of the Princess Takamatsu Cancer Research Fund.

12 February 1985; accepted 29 April 1985

HTLV *x*-Gene Product: Requirement for the *env* Methionine Initiation Codon

Abstract. *The human T-cell leukemia viruses (HTLV) are replication-competent retroviruses whose genomes contain* gag, pol, *and* env *genes as well as a fourth gene, termed* x, *which is believed to be the transforming gene of HTLV. The product of the* x *gene is now shown to be encoded by a 2.1-kilobase messenger RNA derived by splicing of at least two introns. By means of S_1 nuclease mapping of this RNA and nucleic acid sequence analysis of a complementary DNA clone, the complete primary structure of the* x*-gene product has been determined. It is encoded by sequences containing the* env *initiation codon and one nucleotide of the next codon spliced to the major open reading frame of the HTLV-I and HTLV-II* x *gene.*

The human T-cell leukemia viruses (HTLV-I and HTLV-II) are associated with specific T-cell malignancies in man. HTLV-I-related adult T-cell leukemia is endemic to parts of Japan, the Caribbean, and Africa; HTLV-II is associated with a single case of T-cell–variant hairy-cell leukemia (*1–5*). Both viruses will transform normal, human, peripheral blood T cells in vitro as defined by their continued proliferation in the absence of exogenous interleukin-2 (*6–9*). The mechanism of HTLV-induced T-cell transformation is unknown, although these retroviruses, and bovine leukemia virus (BLV), appear to use a mechanism distinct from that of other animal retroviruses. Molecular studies of the HTLV